CODEN: JASMAN 4966

THE JOURNAL of the Acoustical Society of America

Vol. 102, No. 5, Pt. 1 November 1997

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

BIOACOUSTICS [80]

LETTERS TO THE EDITOR

NOTES CONCERNING ARTICLE ABSTRACTS

1. The number following the abstract copyright notice is a Publisher Item Identifier (PII) code that provides a unique and concise identification of each individual published document. This PII number should be included in all document delivery requests for copies of the article.
2. PACS numbers are for subject classification and indexing. See June and December issues for detailed listing of acoustical classes and subclasses.
3. The initials in brackets following the PACS numbers are the initials of the JASA Associate Editor who accepted the paper for publication.

Document Delivery: Copies of articles can be ordered for $15 per copy from the AIP/Member Society document delivery service ''Articles in Physics,'' 75 Varick Street, New York, NY 10013; Fax: 212-301-4060; Telephone: 800-480-PHYS (800-480-7497) (in U.S. and Canada), or 212-301-4000; E-mail: articles@aip.org; URL: http://www.aip.org/articles.html

CONTENTS

REVIEWS OF ACOUSTICAL PATENTS

Daniel W. Martin
7349 Clough Pike, Cincinnati, Ohio 45244

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.

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception Incorporated, Box 39536, Los Angeles, California 90039*
MAHLON D. BURKHARD, *31 Cloverdale Heights, Charles Town, West Virginia 25414*
RONALD B. COLEMAN, *BBN Acoustic Technologies, 70 Fawcett Street, Cambridge, Massachusetts 02138*
HARVEY H. HUBBARD, *325 Charleston Way, Newport News, Virginia 23606*
SAMUEL F. LYBARGER, *101 Oakwood Road, McMurray, Pennsylvania 15317*
D. LLOYD RICE, *11222 Flatiron Drive, Lafayette, Colorado 80026*
WILLIAM THOMPSON, JR., *The Pennsylvania State University, University Park, Pennsylvania 16802*
ERIC E. UNGAR, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, New York 14642*

5,559,759

43.30.Tg METHOD OF ALERTING MARINE MAMMALS AND OTHER MAMMALS UNDERWATER OF THE DANGER OF APPROACHING MOTOR VESSELS

L. A. Gerstein, E. R. Gerstein, Boca Raton, FL and J. E. Blue, Orlando, FL
24 September 1996 (Class 367/139); filed 9 December 1994

A method of alerting manatees, and other underwater mammals, to the danger of an approaching motor vessel consists of radiating a pulsed acoustical signal from a sonar projector (not detailed) in a highly directional manner from the motor vessel and in the direction of travel of the vessel. The signals, in the frequency range from 3 to 26 kHz, are either radiated directly or created parametrically.—WT

5,497,357

43.30.Yj SHOCK-RESISTANT FLEXTENSIONAL TRANSDUCER

D. K. Dahlstrom and T. R. Kazmar, assignors to Allied Signal, Incorporated
5 March 1996 (Class 367/158); filed 23 December 1988

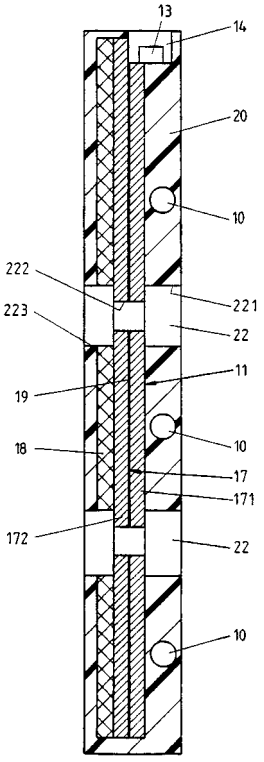
A class IV flextensional transducer is designed to be resistant to damage from an external explosive shock wave by using a shell that is somewhat heavier (i.e., thicker) than usual; by wrapping the cylindrical stacks of piezoceramic disks which constitute the motor units and which are oriented along the major axis of the elliptical cross-section shell with glass fiber cloth to inhibit ceramic chipping, and to put radial precompression on the ceramic; and by positioning adjustable length rods or snubbers along the minor axis of the shell which function to limit the distance that this surface of the shell can deflect inward under the effect of an explosive force.—WT

5,499,219

43.30.Yj ELECTRO-ACOUSTICAL TRANSDUCER ARRANGEMENT

A. Brenner *et al.*, assignors to STN Atlas Elektronik GmbH
12 March 1996 (Class 367/151); filed in Germany 23 November 1993

An underwater electroacoustical transducer array consists of a number (in this case three) of piezoceramic spherical hydrophones **10** equally spaced in a linear array which is intended to be one stave of a cylindrical structure. A mass-spring type of reflector **11** consisting of two metal plates **171** and **172** (the mass element) sandwiching a flexural wave damping layer **19**, and



a nonrigid elastic panel **18** (the spring element) is disposed behind each hydrophone. The electrical leads from all hydrophones are collected in a waterproof electrical connector **13** at one end of the line. The entire assembly is embedded in an acoustically transparent elastomer **20** such as polyurethane.—WT

5,490,220

43.38.Bs SOLID STATE CONDENSER AND MICROPHONE DEVICES

Peter V. Loeppert, assignor to Knowles Electronics, Incorporated
6 February 1996 (Class 381/168); filed 18 March 1992

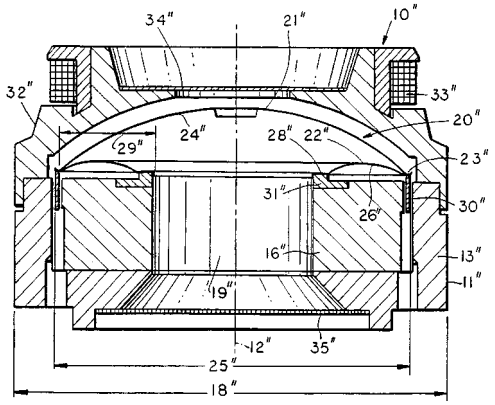
Micro-machining techniques are used to create a condenser microphone from silicon semiconductor materials. By using very small spacing between the back plate and diaphragm, and by supporting the diaphragm with ''keepers'' that minimize diaphragm tension, sensitivities comparable to more conventional condenser microphones are achieved with polarization voltages of only a few volts.—MDB

5,471,437

43.38.Dv ELECTRODYNAMIC ACOUSTIC TRANSDUCER

E. Schutter and V. Gorelik, assignors to Sennheiser Electronic KG
28 November 1995 (Class 367/175); filed in Germany 4 September 1993

The diaphragm of an electrodynamic transducer is realized in two parts. The projected area of one part is a circle, while the projected area of the other part is an annulus. Unlike previous designs where these two parts are arranged so that the outer diameter of the circular part is the inner diameter of the annular part, and the drive coil is positioned at this intermediate diameter, here the outer diameters of the circular diaphragm 20 in. and



annular diaphragm 26 in. are the same. The annular diaphragm is essentially tucked under the circular one, and the drive coil 30 in. is positioned at the outer periphery of both. This results in a smaller overall size for the unit. The reader is cautioned not to conclude from the dimension lines at the bottom of the figure that 18 in. is now greater than 25 in.—WT

5,486,734

43.38.Fx ACOUSTIC TRANSDUCER USING PHASE SHIFT INTERFERENCE

Mir S. Seyed-Bolorforosh, Palo Alto, CA
23 January 1996 (Class 310/327); filed 18 February 1994

An array of delay sections is created on the back or front radiating surface of a piezoelectric element. The delay sections are intended to control the shape of emitted pulses and, in particular, to reduce ringdown or decay time.—MDB

5,493,541

43.38.Fx ULTRASONIC TRANSDUCER ARRAY HAVING LASER-DRILLED VIAS FOR ELECTRICAL CONNECTION OF ELECTRODES

J. E. Snyder, assignor to General Electric Company
20 February 1996 (Class 367/155); filed 30 December 1994

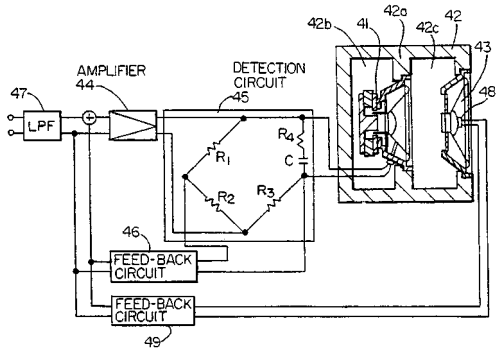
A flat plate of piezoelectric material is drilled through its thickness dimension at a number of positions by a high-powered laser. When these vias are sputtered or plated with electrically conducting material, electrical connections can then be made, at the rear of the plate, between electrodes on the front face of the piezoelectric plate and a signal source or a common ground.—WT

5,588,065

43.38.Ja BASS REPRODUCTION SPEAKER APPARATUS

Shoji Tanaka *et al.*, assignors to Matsushita Electric Industrial Company
24 December 1996 (Class 381/96); filed in Japan 20 December 1991

The goal is to combine a small bandpass loudspeaker with a power amplifier in such a way that it behaves like a larger system tuned to a lower frequency. In one possible design (ten are included in the patent) a passive radiator is used in place of the more conventional vent. This allows motional



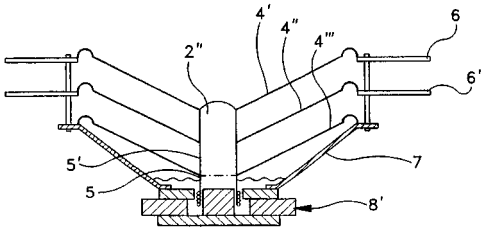
feedback to be derived from the passive radiator. Current feedback is taken from the driving loudspeaker. The two are combined through some fancy, computer-derived circuitry. Voila, a miniature infra-bass system (the patent actually says ''ultra bass band'').—GLA

5,594,804

43.38.Ja MULTIPLE CONE TYPE LOUDSPEAKER

Kyung W. Kim, Choonchun City, Republic of Korea
14 January 1997 (Class 381/186); filed in Korea 29 April 1994

Loudspeaker **7** is actually mounted in a conventional box. Two additional cones **4′**, **4″** and baffle plates **6**, **6′** project in front of the speaker mounting panel. ''Accordingly, it is possible to achieve an extension [of the]



regeneration band, an improvement in directivity and a generation of dynamic sound.''—GLA

5,615,275

43.38.Ja PLANAR DIAPHRAGM LOUDSPEAKER WITH COUNTERACTIVE WEIGHTS

Eduardo J. Bertagni, assignor to Sound Advance Systems, Incorporated
25 March 1997 (Class 381/203); filed 21 April 1995

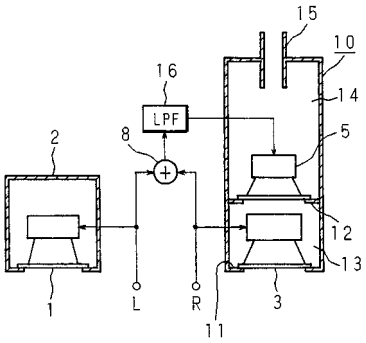
It is an old technique to suppress loudspeaker cone resonances by adding blobs of goop at specific locations. In the case of molded planar diaphragms, this patent establishes that resiliently mounted weights can serve a similar function and can be incorporated into the design of the diaphragm.—GLA

5,621,804

43.38.Ja COMPOSITE LOUDSPEAKER APPARATUS AND DRIVING METHOD THEREOF

Satoshi Beppu, assignor to Mitsubishi Denki Kabushiki Kaisha
15 April 1997 (Class 381/90); filed in Japan 28 December 1993

What is shown is one of 16 possible versions, all described at length in the patent document. Loudspeakers 1 and 3 are identical, full-range units. At very low frequencies, loudspeakers 3 and 5 function as a composite unit to



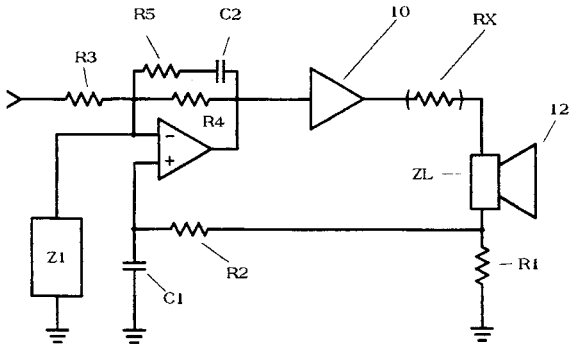
drive vented chamber 14. So long as the bass content of left and right channels is not too different, the arrangement will work.—GLA

5,625,698

43.38.Ja LOUDSPEAKER AND DESIGN METHODOLOGY

Anthony T. Barbetta, Moorpark, CA
29 April 1997 (Class 381/96); filed 14 August 1995

Previously it has been demonstrated that an overdamped loudspeaker/box combination can be realigned by using just the right amount of negative current feedback. The converse is also true, and this is the approach taken in



the patent. According to the patent, the circuitry shown provides stability over a wider bandwidth than previous designs.—GLA

5,490,219

43.38.Kb NOISE CANCELING MICROPHONE WITH VISUAL FEEDBACK

Mehrdad Badie *et al.*, assignors to Motorola, Incorporated
6 February 1996 (Class 381/122); filed 4 March 1991

''The microphone comprises a sound transducer for converting sound waves to electrical signals and means for substantially preventing ambient noise from being converted to electrical signal. A circuit is provided for determining the level of noise not prevented from being converted to electrical signal.''—MDB

5,579,404

43.38.Lc DIGITAL AUDIO LIMITER

Louis D. Fielder *et al.*, assignors to Dolby Laboratories Licensing Corporation
26 November 1996 (Class 381/106); filed 16 February 1993

In the wonderful world of perceptual encoding and digital transforms, it turns out that the process of encoding a peak-limited audio signal can preserve apparent loudness but lose peak limiting. The patent is full of information about this new technology and presents a practical way of correcting the problem.—GLA

5,640,458

43.38.Md AUDIO SIGNAL REPRODUCING APPARATUS

Masayuki Nishiguchi and Yoshihito Fujiwara, assignors to Sony Corporation
17 June 1997 (Class 381/74); filed in Japan 25 October 1989

The patent shows a headset with earphones held on each ear by a headband. Adjacent to the earphones are compartments containing the battery operated portions of the device. Digitized and compression encoded audio signals stored in a semiconductor memory are read out so as to undergo a decoding operation to reproduce the audio signals and deliver them to the earphones. A block diagram of the device is shown and the method of encoding described. The system permits high-fidelity audio signals to be stored that on playback last 4 min on a 16 M bit ROM chip. Four such chips provide a playback time of about 17 min. With improved semiconductor memories, playback times of 33 or 66 min may be possible.—SFL

5,617,158

43.38.Ne DIGITAL SOUND RECORDING ON MOTION PICTURE FILM

Shinji Miyamori *et al.*, assignors to Sony Corporation and Sony Cinema Products Corporation
1 April 1997 (Class 352/37); filed in Japan 7 December 1993

Sony's method of optically recording digital sound tracks is augmented to allow hand-cut editing while maintaining adequate error correction. The patent is clearly written and includes more than a dozen helpful illustrations.—GLA

5,483,599

43.38.Tj DIRECTIONAL MICROPHONE SYSTEM

Michael A. Zagorski, St. John's, Newfoundland, Canada
9 January 1996 (Class 381/68.5); filed 28 May 1992

''A directional microphone system gain adjusts the sum of signals that are received from two microphones mounted spaced apart. The gain adjustment is generally proportional to the product of the two microphone signals. Analog multipliers are used both for multiplying the signals together and also to form a gate that gain adjusts the sum of the signals.''—MDB

5,572,591

43.38.Vk SOUND FIELD CONTROLLER

Hiroko Numazu *et al.*, assignors to Matsushita Electric Industrial Company
5 November 1996 (Class 381/1); filed in Japan 9 March 1993

By now there must be more than a dozen patents describing sound field controllers. This patent document is unusual in that it seems to be written in single-sentence paragraphs, each more than 100 words in length. The controller attempts to envelop a listener in panoramic, reverberant sound without degrading the primary stereophonic signal.—GLA

5,602,923

43.38.Vk THEATER SOUND SYSTEM WITH UPPER SURROUND CHANNELS

Yoshio Ozaki and Michael J. Kohut, assignors to Sony Corporation and Sony Cinema Products Corporation
11 February 1997 (Class 381/18); filed 23 August 1995

Pilot signals are used to redirect surround channels to overhead loudspeakers for special effects.—GLA

5,563,845

43.40.Ph SYSTEM AND METHOD FOR ACOUSTICALLY IMAGING AN UNDERGROUND TANK

K. M. Walsh, assignor to the United States of America
8 October 1996 (Class 367/7); filed 7 November 1995

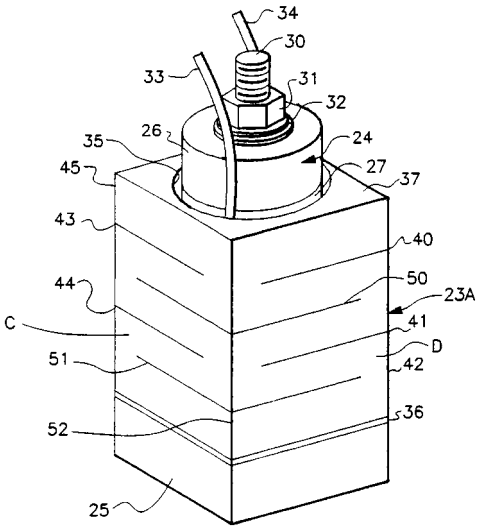
This patent discusses a vertically oriented line array and associated electronics for acoustically imaging the contents of a buried tank. The partially filled tank may contain radioactive wastes that have stratified into horizontal layers, some of which may contain gas bubbles. The transducer array is placed in a water-filled vertical tube located near the edge of the buried tank. The dirt located between the tube and the tank is saturated with water to facilitate acoustic coupling between the transducer array and the tank. The array consists of one or more transmitting and receiving arrays. The receiving elements include dipoles to produce some directionality in the horizontal plane. The associated topside electronics unit produces a variety of signals to the transmitters to aid in obtaining useful acoustic images.—WT

5,491,671

43.40.Sk SONAR TRANSDUCER WITH UNITARY ISOLATOR

A. F. Potter, assignor to Alliant Techsystems, Incorporated
13 February 1996 (Class 367/158); filed 27 November 1987

A vibration isolator **23A** for a head-mounted Tonpilz-type transducer **24** comprises a block of material such as beryllium copper with a central cavity **27** to accommodate the transducer and into which a number of transverse slits **40**, **41**, **43**, **44**, **50**, and **51** have been cut. The cross-sectional dimensions of the block match those of the head mass **25** of the transducer, which in the case shown is square. These slits extend inwardly but stop short of the longitudinal axis of the block. The pair of slits **40** and **43** extend into



the block from diagonally opposite corners **42** and **45** as do the next pair, **50** and one not seen in the figure, that are axially spaced from the previous pair; this latter pair originate at the other two corners of the block. This pattern is repeated along the entire axial length of the block. The structure behaves as a linear spring up to hydrostatic pressure levels where the narrow slits become closed.—WT

5,624,104

43.40.Tm VIBRATION ISOLATING SUPPORTER

Koumei Hukuda *et al.*, assignors to Nippon Steel Corporation and Showa Electric Wire and Cable Company
29 April 1997 (Class 267/34); filed in Japan 16 January 1996

This vibration-isolating support for machinery employs steel coil springs, arranged in mechanical parallel between a base plate and a top support plate. Except for a short portion near its top, each spring is immersed in a damping material, such as silicone gel, contained between suitable walls. A volume-displacing element extending along the axis of each spring is attached to the top plate, so as to enhance the damping material's pumping through the coils as the top plate vibrates vertically relative to the base plate.—EEU

5,627,425

43.40.Tm VIBRATING UNIT

Hiroaki Kaida and Jiro Inoue, assignors to Murata Manufacturing Company
6 May 1997 (Class 7310/321); filed in Japan 3 July 1992

This patent relates to a resonant piezoelectric oscillator to which there are connected one or more dynamic absorbers (or ''tuned dampers''). These absorbers, which consist of appropriately designed flexural members—such as beams or plates—are intended to reduce the vibrations transmitted to the oscillator's supports at the oscillator's resonance frequencies. Configurations of ''chip'' oscillators with such dynamic absorbers are illustrated.—EEU

5,564,537

43.40.Vn ADAPTIVE-PASSIVE VIBRATION CONTROL SYSTEM

Rahmat A. Shoureshi, assignor to Cooper Tire and Rubber Company
15 October 1996 (Class 188/378); filed 4 April 1994

A system for actively tuning passive resonant absorbers is discussed for reducing vibration transmission between structures. Either the mass or the stiffness of the tuned absorbers is adjusted to maintain the resonance

frequency of the absorber at the desired frequency. An accelerometer is used to sense the frequency to be controlled. A controller is then used to adjust the mass of a fluid cavity, or the volume of an air spring, to modify the mass and stiffness of a tuned absorber, respectively.—RBC

5,578,881

43.40.Vn AXIAL VIBRATION DAMPING ARRANGEMENT

Michael K. Swann, assignor to Glacier RPB, Incorporated
26 November 1996 (Class 310/90.5); filed 29 September 1994

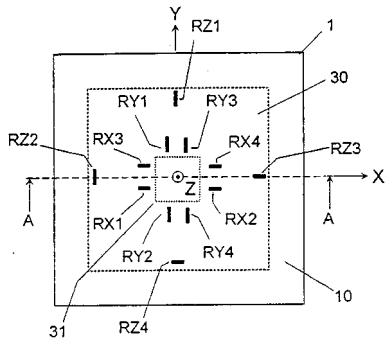
This patent describes a system for damping axial vibrations in turbomachine shafts. The system can be used to accommodate the failure of magnetic bearings. The system measures pressure fluctuations in cavities located on either side of a thrust collar mounted on the shaft. These pressure fluctuations serve as inputs to a controller that pulses gas pressure in the cavities. The phase relationship between the sensed and actuated pressures is adjusted to provide axial damping.—RBC

5,485,749

43.40.Yq ACCELERATION DETECTOR

Kazuya Nohara and Naohiro Taniguchi, assignors to Matsushita Electric Works
23 January 1996 (Class 73/517.34); filed in Japan 26 May 1993

An accelerometer is fabricated by micro-machining silicon. Strain within the structure is sensed by resistance changes within piezoresistance



elements **RX1**–**RZ4**. Resistance elements are arranged such that combinations enable detection of acceleration along three orthogonal axes.—MDB

5,610,837

43.40.Yq SYSTEM AND METHOD FOR NONDESTRUCTIVE VIBRATIONAL TESTING

Joseph F. Murphy, assignor to Sonoco Products Company
11 March 1997 (Class 364/508); filed 21 April 1994

This patent relates to nondestructive testing of specimens with a defined geometry, such as paper core tubes. The test specimen is supported on a pair of knife edges, under one of which there is a load cell which serves as a vibration sensor. The system measures the specimen's natural frequencies and decay constants, and from these calculates the dynamic elastic moduli and damping values. The data acquisition components interface with a conventional personal computer, which does all of the necessary data processing.—EEU

5,618,989

43.40.Yq ACCELERATION SENSOR AND MEASUREMENT METHOD

Jiri Marek, assignor to Robert Bosch GmbH
8 April 1997 (Class 73/1.38); filed in Germany 15 September 1994

The seismic mass of this accelerometer is supported via soft flexures above a planar substrate, so that it can move essentially parallel to the substrate's surface as the substrate is subjected to acceleration parallel to that surface. The seismic mass is provided with several electrodes that face matching electrodes which are rigidly connected to the substrate. In one type of implementation, the capacitances between electrode pairs are measured to obtain an indication of the displacement of the mass relative to the substrate. In another type of implementation, some electrode pairs are used to provide a force that keeps the mass in its initial equilibrium position, and the voltage required to provide this force is measured as an indication of the acceleration. Application of a known voltage between one of the electrode pairs can provide a known force on the mass, thus making it possible to determine whether the accelerometer is working properly.—EEU

5,531,287

43.50.Gf POWER STEERING SYSTEM WITH VIBRATION AND NOISE SUPPRESSION

Howard M. Sherman, assignor to Chrysler Corporation
2 July 1996 (Class 180/417); filed 27 March 1995

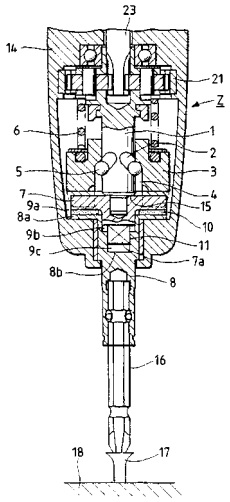
This patent relates to the control of rack rattle in the conventional power steering systems of automotive vehicles. This is accomplished by providing a flow restriction device in the return line, preferably located a substantial distance from the control valve, to increase the back pressure and place a restraint on the return flow.—HHH

5,601,149

43.50.Gf NOISE REDUCTION MECHANISM FOR PERCUSSION TOOLS

Yoshimitsu Kawasaki *et al.*, assignors to Hitachi Koki Company
11 February 1997 (Class 173/93.5); filed in Japan 25 February 1994

This patent relates to the control of noise from handheld percussion tools. A main source of noise, which results from the axial forces of the



spindle and hammer against the anvil, can be reduced by inserting an elastic cushion between the spindle and anvil.—HHH

5,601,410

43.50.Gf FAN HAVING BLADES WITH SOUND REDUCING MATERIAL ATTACHED

Daniel A. Quinlan, assignor to Lucent Technologies, Incorporated
11 February 1997 (Class 416/241A); filed 31 August 1995

This patent relates to the control of noise from small, low-speed fans of the type used for air cooling of electronic systems. Materials such as felt, batting, and hook and loop (velcro) type fasteners, which have preapplied adhesives, are attached as thin strips along the trailing edges of the fan blades. The above materials, which in all cases are less dense than the fan blade material, are noted to supply some sound absorption and to alter the aerodynamic flow beneficially.—HHH

5,605,447

43.50.Gf NOISE REDUCTION IN A HERMETIC ROTARY COMPRESSOR

Han-Jun Kim and In-Soo Hwang, assignors to Carrier Corporation
25 February 1997 (Class 417/312); filed 3 July 1996

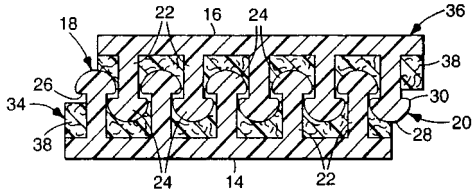
This patent relates to the control of low-frequency noise generated in a hermetic rotary compressor. Muffler outlets are separated in distance by 1/4 to 1/2 the wavelength of the principal pulsating flow frequency in order to provide noise reduction by destructive interference.—HHH

5,611,122

43.50.Gf INTERENGAGING FASTENER HAVING REDUCED NOISE GENERATION

Shinji Torigoe *et al.*, assignors to Minnesota Mining and Manufacturing Company
18 March 1997 (Class 24/442); filed in Japan 28 July 1993

This patent relates to an interengaging fastener that is said to be quieter in operation than conventional interengagement fasteners. Noise reduction



arises from shock absorbing material such as glass wool, sponge, or nonwoven fabric, which is disposed around the stems of the interlocking headed elements to restrict their relative motions.—HHH

5,611,301

43.50.Gf STRUCTURAL ENCLOSURE OF COMBUSTION ENGINES FOR THE PURPOSE OF REDUCING ENGINE NOISE

Per Gillbrand and Lars Bergsten, assignors to SAAB Automobile Aktiebolag
18 March 1997 (Class 123/48C); filed in Sweden 28 May 1993

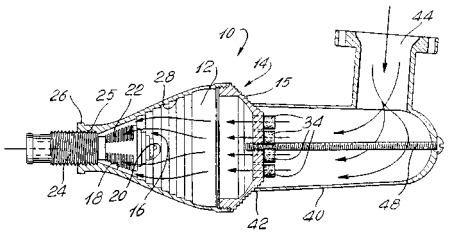
The patent relates to the reduction of noise of internal combustion engines used in automotive vehicles. Included are considerations of using a lateral wall spaced outward from the cylinder section and the choice of materials for enhancing thermal conductivity while damping vibration and noise.—HHH

5,611,409

43.50.Gf EXHAUST MUFFLER FOR SMALL INTERNAL COMBUSTION ENGINE

Michel Arseneau, St. Timothée, Quebec, Canada
18 March 1997 (Class 181/228); filed 9 May 1995

The patent relates to light-weight exhaust mufflers for two-cycle engines of the type used on small model airplanes. An expansion chamber is



provided, along with tube-shaped neoprene absorbers for acoustic absorption, and multiple slits for regulating backpressure.—HHH

5,499,302

43.50.Ki NOISE CONTROLLER

Masaaki Nagami *et al.*, assignors to Fujitsu Ten Limited
12 March 1996 (Class 381/71); filed in Japan 26 May 1992

This patent describes an adaptive feedback system to control the interior noise in an automobile. The loudspeakers are used to radiate the canceling signals, and microphones near the occupant's head positions are used to determine the responses to be minimized. The control filter attempts to invert (or equalize) the plant transfer function between the drive signals to the loudspeakers and the microphones. The control filter is adapted using a filtered-X least-mean squares (FX-LMS) algorithm in conjunction with a white noise probe signal. Functionality is included to judge the accuracy of the plant equalization and to turn off the controller in the event of changes in the plant transfer function. In that case, the controller is re-initialized before continuing closed-loop operation.—RBC

5,519,637

43.50.Ki WAVENUMBER-ADAPTIVE CONTROL OF SOUND RADIATION FROM STRUCTURES USING A ''VIRTUAL'' MICROPHONE ARRAY METHOD

Gopal Mathur, assignor to McDonnell Douglas Corporation
21 May 1996 (Class 364/508); filed 20 August 1993

This patent describes a method for controlling sound radiation from and noise transmission through a structure. The approach is primarily focused at reducing radiation from elastic structures such as aircraft cabin panels. The control algorithm uses arrays of accelerometers on the panel to characterize the vibration response. Wave number–frequency transforms of the vibration response are then taken to determine the radiating (i.e., acoustically fast) response on the panel as a function of frequency. These radiating components are then projected to a plane in the far field using the forward propagation equation of near-field acoustic holography. A least-mean-squares (LMS) algorithm is then used to determine drive signals to actuators on the panel to reduce the estimated far-field radiated pressure.—RBC

5,525,853

43.50.Ki SMART STRUCTURES FOR VIBRATION SUPPRESSION

Theodore W. Nye *et al.*, assignors to TRW, Incorporated
11 June 1996 (Class 310/316); filed 2 June 1994

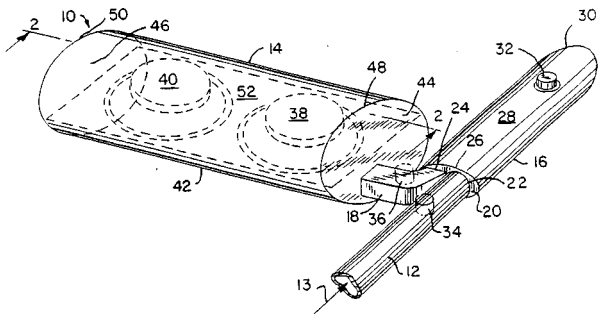
Modular control patches, including assemblies of sensors and actuators, are provided to impart or control vibration of foundation members such as struts in aerospace apparatus. The patches may be chemically bonded to the outside of the struts or imbedded within a composite structure. Each patch is controlled as a single-input–single-output system. The controller described uses analog electronics to provide active damping at each patch location. Coupling between patches is not discussed.—RBC

5,541,373

43.50.Ki ACTIVE EXHAUST SILENCER

C. Raymond Cheng, assignor to Digisonix, Incorporated
30 July 1996 (Class 181/206); filed 6 September 1994

An active silencer for a vehicle exhaust system is described that has a low profile so that it can be hung underneath the vehicle in a conventional manner without altering the vehicle. The silencer has a chamber **14** isolated from the exhaust flow **13** in which one or more loudspeakers are mounted side-by-side. The loudspeakers **38**, **40** face downward and share a common



tuning chamber within the isolated chamber. The tuning chamber is ported **18** to the mixing chamber **16** where the canceling acoustic waves mix with noise from the exhaust flow before the exhaust noise exits to the atmosphere **30**.—RBC

5,546,467

43.50.Ki ACTIVE NOISE ATTENUATED DSP UNIT

Jeffrey N. Denenberg, assignor to Noise Cancellation Technologies, Incorporated
13 August 1996 (Class 181/206); filed 14 March 1994

This is an interesting patent aimed at using active control to reduce the noise radiated from one or more home appliances. Rather than imbedding the control electronics in the appliances (e.g., hair blower), the approach is to locate the digital-signal-processing (DSP) unit at the ac power source and connect it to the anti-noise actuators and sensors inside the appliance via signals sent and received over the ac power line.—RBC

5,548,653

43.50.Ki ACTIVE CONTROL OF NOISE AND VIBRATIONS IN MAGNETIC RESONANCE IMAGING SYSTEMS USING VIBRATIONAL INPUTS

Frederic C. Pla *et al.*, assignors to General Electric Company
20 August 1996 (Class 381/71); filed 20 August 1993

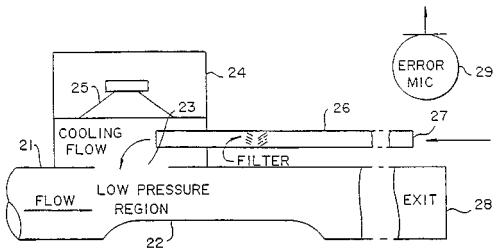
An active control system is described that minimizes noise and vibration associated with magnetic resonance imaging. The system includes one or more piezoceramic actuators mounted to the inner surface of the imaging device. Transducers provide for sensing the noise or vibrations generated by the device, and produce an error signal corresponding to the level of noise or vibration sensed. The system is nominally configured as a feedforward LMS controller using the gradient pulse signals of the device as a reference signal. This system is said not only to reduce the noise levels generated by the device, which approach 100 dB, but also to maintain good image quality.—RBC

5,550,334

43.50.Ki ACTIVELY SOUND REDUCED MUFFLER HAVING A VENTURI EFFECT CONFIGURATION

Andrew J. Langley, assignor to Noise Cancellation Technologies, Incorporated
13 August 1996 (Class 181/206); filed 14 March 1994

This patent describes an active muffler system designed to minimize the acoustic mass between the loudspeaker and the error microphone, and to minimize the ingestion of hot exhaust gases into the loudspeaker enclosure. The acoustic mass is kept small by mounting the loudspeaker enclosure **24** directly to the muffler pipe **21**. A smooth venturi **22** is used to generate



locally in the exhaust pipe a region that is below atmospheric pressure. A bleed pipe **26** in the front cavity of the loudspeaker enclosure permits air from outside to flow toward the low-pressure region, thereby reducing ingestion of exhaust gases into the loudspeaker enclosure, and also providing cooling.—RBC

5,553,153

43.50.Ki METHOD AND SYSTEM FOR ON-LINE SYSTEM IDENTIFICATION

Graham P. Eatwell, assignor to Noise Cancellation Technologies, Incorporated
3 September 1996 (Class 181/206); filed 10 February 1994

This patent relates to an improved method of on-line system identification for use with active control systems that requires less computation and reduces the problem of coefficient jitter in the filters. The approach uses a circulant, block-white probe signal, which results in less computations per filter coefficient than traditional LMS-based algorithms which use random noise probes. In addition, the effect of the probe signal is subtracted from the residual sensor response, which reduces coefficient jitter.—RBC

5,504,718

43.58.Ls BEAM CONVERGENCE MODULE

R. E. Gaynor, assignor to Wadi, Incorporated
2 April 1996 (Class 367/138); filed 17 October 1994

A length of honeycomb material (hexagonal cross-section free-flooding pipes), placed with the pipe axes perpendicular to the radiating face of a transducer, acts as a lens and significantly decreases the beamwidth of the directivity pattern. A similar effect is obtained for a microwave radar device.—WT

5,603,726

43.66.Ts MULTICHANNEL COCHLEAR IMPLANT SYSTEM INCLUDING WEARABLE SPEECH PROCESSOR

Joseph H. Schulman *et al.*, assignors to Alfred E. Mann Foundation For Scientific Research
18 February 1997 (Class 607/57); filed 22 September 1989

A cochlear implant hearing aid is described for which part of the signal processing functions are implemented in a wearable, external computer pack, and part are included in a small, subdermally implanted electronics unit. The external computer receives the microphone signal and performs gain control and A-to-D conversion. A filter bank spectrum analyzer is implemented either before or after the A/D step. Most of the patent is concerned with details of the telemetry of the spectral information by radio link to the implanted processor, which controls 16 current source electrodes embedded in the cochlea. Electrode activation uses a form of pulse width modulation through log scale D/A converters.—DLR

5,633,937

43.66.Ts METHOD FOR PROCESSING SIGNALS

Zlatan Ribic, assignor to Viennatone AG
27 May 1997 (Class 381/68.4); filed in Austria 13 November 1991

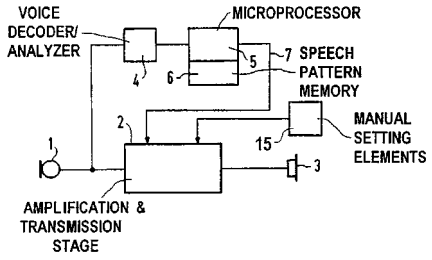
In a hearing aid amplifier, a delay-free processing of signals is achieved in that a momentary amplitude signal $A(t)$ is produced having its strength proportional to the momentary amplitude of the input signal $u(t)$, and that the momentary amplitude signal $A(t)$ is used for processing the input signal. The method eliminates the undesired delay in response behavior which can occur, for example, when very weak sounds occur directly after loud sounds. The circuitry involves the production of Hilbert transforms from the input signal.—SFL

5,636,285

43.66.Ts VOICE-CONTROLLED HEARING AID

Joseph Sauer, assignor to Siemens Audiologsiche Tecknik GmbH
3 June 1997 (Class 381/68.2); filed in Germany 7 June 1994

The hearing aid described is provided with a microprocessor-controlled speech pattern memory in parallel with the hearing aid amplification circuit. Code words, such as ''softer'' or ''louder,'' spoken by the wearer, are stored in the memory and are programmed to change the amplifier gain as requested. Other stored commands might be ''on'' or ''off'' or ''program 1.'' The voice control method would be particularly valuable for extremely small hearing aids. A modification of the system using fuzzy logic, and another using a neural network, are also shown.—SFL

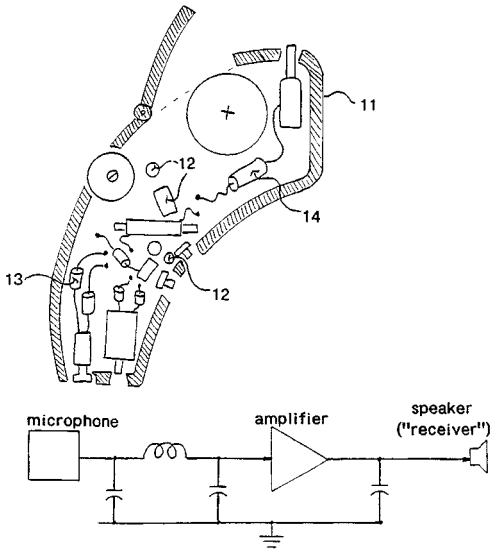


5,640,457

43.66.Ts ELECTROMAGNETICALLY SHIELDED HEARING AID

Louis Thomas Gnecco and Paula Sharyn Gnecco, Herndon, VA
17 June 1997 (Class 381/69); filed 13 November 1995

Shielding of hearing aids to make them resistant to interference produced by modular telephones in the 800—1000 MHz frequency range is described. In addition to electrostatic shielding in the hearing aid case



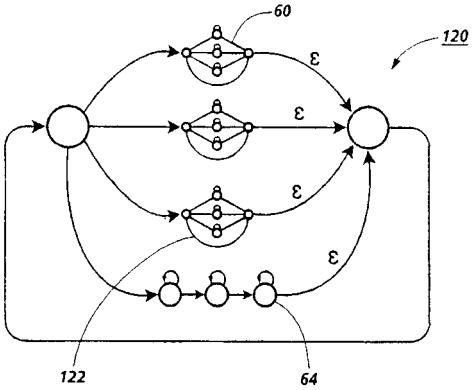
(this dates back to about 1938), shunt capacitance and series inductance elements effective at the cellular telephone frequencies are added at strategic locations.—SFL

5,598,507

43.72.Ar METHOD OF SPEAKER CLUSTERING FOR UNKNOWN SPEAKERS IN CONVERSATIONAL AUDIO DATA

Donald G. Kimber *et al.*, assignors to Xerox Corporation
28 January 1997 (Class 395/2.55); filed 12 April 1994

This patent describes a method for determining similarities between speaking voices as a part of a speaker identification system. For purposes of computing distances between speakers, the speech signal is analyzed into features, as for example, 12th-order cepstral vectors computed every 20 ms.



A method known as hierarchical agglomerative clustering is used to construct clusters in the feature space which represent the individual speakers.—DLR

5,608,839

43.72.Ar SOUND-SYNCHRONIZED VIDEO SYSTEM

Homer H. Chen, assignor to Lucent Technologies, Incorporated
4 March 1997 (Class 395/2.44); filed 18 March 1994

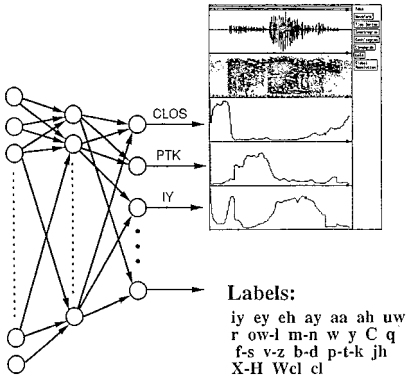
This device correlates spectral features of a speech signal with visual features extracted from a video image of the speaker's mouth to synchronize the sound track with the video display. The video image is compared to a facial model and, in particular, with a mouth model such that the moving mouth parts can be tracked. A mouth deformation module classifies mouth movements into phonetically relevant units by reference to a lookup table of 12 movement elements referred to as visemes. The sequence of mouth movement units is then directly mappable into their acoustic correlates for matching with the spectral data.—DLR

5,621,857

43.72.Bs METHOD AND SYSTEM FOR IDENTIFYING AND RECOGNIZING SPEECH

Ronald A. Cole and Mark A. Fanty, assignors to Oregon Graduate Institute of Science and Technology
15 April 1997 (Class 395/2.41); filed 20 December 1991

This speech recognition system uses a neural network component to recognize telephone speech consisting of spoken and spelled proper names. Either DFT or perceptual linear prediction spectra of 10-ms frames are computed along with peak amplitude and zero-crossing counts every 3 ms. The feature vectors are then passed to a three-layer, feed-forward, neural net



which assigns a score for each of twenty two phonetic categories. A modified Viterbi search having reference to a directory containing a list of the candidate names then finds the most likely sequences of letter models.—DLR

5,602,962

43.72.Dv MOBILE RADIO SET COMPRISING A SPEECH PROCESSING ARRANGEMENT

Walter Kellermann, assignor to U.S. Philips Corporation
11 February 1997 (Class 395/2.35); filed in Germany 7 September 1993

This patent describes a speech processor which operates on multichannel inputs from an array of two or more microphones to produce an enhanced speech signal with reduced background noise. Each microphone signal is first delayed by an adapting interval to maximally align all of the speech components. The delay adaptation method is not described. An FFT spectrum computed for each signal is then subdivided into two or more subbands. The frequency bands from each microphone are weighted and summed and passed through an adaptive Weiner filter to produce the speech output. Standard deviations of the delayed signals and the individual frequency bands are used to estimate signal-to-noise ratios from which the band weights are adjusted.—DLR

5,610,991

43.72.Dv NOISE REDUCTION SYSTEM AND DEVICE, AND A MOBILE RADIO STATION

Cornelis P. Janse, assignor to U.S. Philips Corporation
11 March 1997 (Class 381/92); filed in European Patent Office 6 December 1993

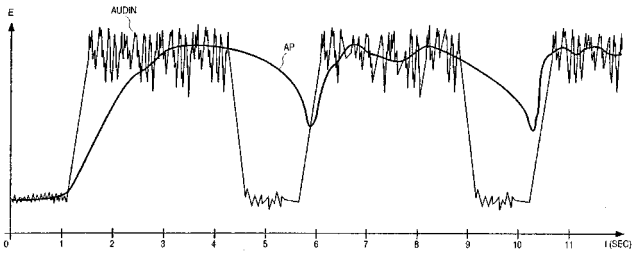
This microphone array processor combines two previously known techniques to avoid the disadvantages of each. In the Zelinski method, the microphone signals are individually delayed, transformed, and Wiener filtered, using auto- and cross-power spectra to control the filter weights. This system has poor frequency characteristics if the microphones are close together. In the Kroschel system, the microphone signals are transformed, delayed, and then sums and differences are formed in the frequency domain. This system is computationally expensive and introduces an overall signal delay equal to the transform size. The method disclosed in this patent includes a speech presence detector so that signal-to-noise ratio estimates can be made independently of the cross-power estimates.—DLR

5,598,466

43.72.Kb VOICE ACTIVITY DETECTOR FOR HALF-DUPLEX AUDIO COMMUNICATION SYSTEM

David L. Graumann, assignor to Intel Corporation
28 January 1997 (Class 379/389); filed 28 August 1995

In this voice detector, the signal energy and standard deviation within 20-ms windows are computed every 20 ms. The energy is then averaged over five windows and the maximum of the averages over the last 2 s is the average peak value (''AP'' in the figure). If the current frame energy is 9 dB or more below the average peak, a noise floor update is considered. If the



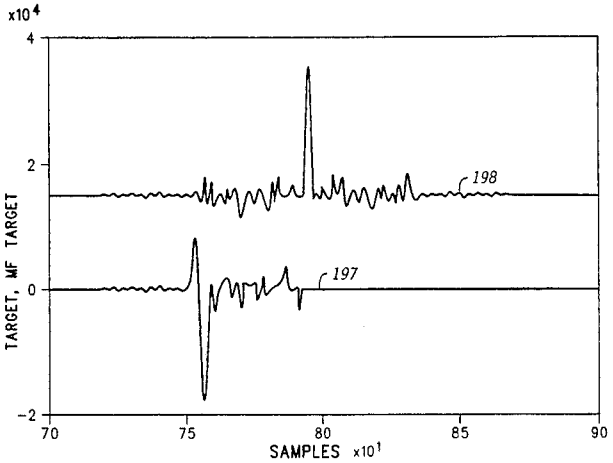
standard deviation has remained at less than 3.2 for at least 1.5 s, the noise floor is updated with a high confidence factor. The same low standard deviation for a shorter time gives a lower confidence factor. Speech presence is indicated if the current energy is greater than the noise floor.—DLR

5,602,959

43.72.Gy METHOD AND APPARATUS FOR CHARACTERIZATION AND RECONSTRUCTION OF SPEECH EXCITATION WAVEFORMS

Chad S. Bergstrom *et al.*, assignors to Motorola, Incorporated
11 February 1997 (Class 395/2.14); filed 5 December 1994

This code excited LPC vocoder is said to offer improved quality of the transmitted voice while maintaining a low coding bitrate. The approach used is to minimize the spectral phase slope and its variance using a pitch-epoch-synchronous coding method. A secondary correction consists of a reduction



of the error resulting from the initial excitation estimate. A continuous envelope reconstruction avoids errors related to prior-art piecewise reconstruction. Spectral group delay is also removed from the excitations.—DLR

5,602,961

43.72.Gy METHOD AND APPARATUS FOR SPEECH COMPRESSION USING MULTI-MODE CODE EXCITED LINEAR PREDICTIVE CODING

Victor D. Kolesnik *et al.*, assignors to Alaris, Incorporated
11 February 1997 (Class 395/2.32); filed 31 May 1994

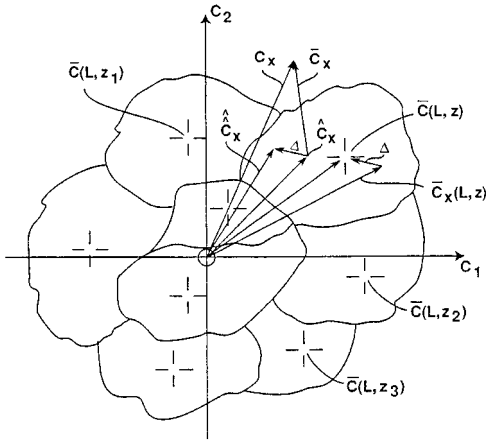
This code-excited (CELP) coder is said to provide high-quality speech at rates as low as 2400 bits per second. Linear prediction coefficients for each speech frame are converted to line spectral pairs coefficients to reduce the bitrate. To code the excitation, each speech frame is divided into from two to six subframes, each of which may use one of three available excitation strategies. Excitations may consist of a train of shaped pulses, or an adaptive codebook search, or a combination of a short-form adaptive search and a stochastic codebook search may be used.—DLR

5,598,505

43.72.Ne CEPSTRAL CORRECTION VECTOR QUANTIZER FOR SPEECH RECOGNITION

Stephen C. Austin and Adam B. Fineberg, assignors to Apple Computer, Incorporated
28 January 1997 (Class 395/2.35); filed 30 September 1994

This speech recognizer uses a vector quantization codebook of cepstral coefficients to reduce the effect of environmental noise. During a training phase, the codebook is constructed by clustering cepstral coefficient sets from speech in a noise-free environment or in a variety of different noise environments. During operation in the recognition environment, an average nonspeech cepstral vector is collected. As a coarse first step, this average is subtracted from subsequent cepstral vectors. The difference vector is then looked up in the cepstral codebook, which yields a fine-tuning adjustment to the cepstral vector.—DLR



5,602,960

43.72.Ne CONTINUOUS MANDARIN CHINESE SPEECH RECOGNITION SYSTEM HAVING AN INTEGRATED TONE CLASSIFIER

Hsiao-Wuen Hon *et al.*, assignors to Apple Computer, Incorporated
11 February 1997 (Class 395/2.16); filed 30 September 1994

This recognizer for Mandarin Chinese speech includes a special, integrated system for analyzing and classifying the tone patterns of spoken Chinese. Whereas prior tone pattern classifiers have often used a gross approximation, such as the overall average pitch slope, this system first performs a syllable segmentation and then computes pitch estimates for each syllable. The sequence of syllable pitch estimates is compared to stored tone patterns for the language, and a set of *N*-best tone theories results from each search. Overall, a long-term tone confidence measure rates how well the sequence of tone theories fits the language structure.—DLR

5,606,645

43.72.Ne SPEECH PATTERN RECOGNITION APPARATUS UTILIZING MULTIPLE INDEPENDENT SEQUENCES OF PHONETIC SEGMENTS

Hiroshi Matsuura, assignor to Kabushiki Kaisha Toshiba
25 February 1997 (Class 395/2.65); filed 28 February 1992

This speech recognizer computes mel-scale cepstral coefficients from linear prediction analyses of the input speech data frames. A phonetic sequence dictionary represents sequences of phonetic units in terms of sequences of the mel-scale feature vectors. A hidden Markov model recognizer finds best matching phonetic sequences for an input utterance. From the matching phonetic sequences, a list of the best scoring words is generated.—DLR

5,613,037

43.72.Ne REJECTION OF NON-DIGIT STRINGS FOR CONNECTED DIGIT SPEECH RECOGNITION

Rafid A. Sukkar, assignor to Lucent Technologies, Incorporated
18 March 1997 (Class 395/2.65); filed 21 December 1993

This speech recognizer achieves high performance on sequences of continuously spoken digits by combining a digit model with a filler model for each digit to improve the rejection of nondigit elements. Feature vectors consist of LPC cepstra, delta and delta–delta cepstra, energy, delta and delta–delta energy. A hidden Markov model recognizer generates a confidence score for each digit in a sequence. If any of the digit confidence scores falls below a digit-specific threshold, then the entire sequence is rejected.—DLR

5,639,976

43.75.Kk CYMBAL FOR PERCUSSION MUSICAL INSTRUMENTS

Hector Ariel Manoff and Gaston Costa Sanjurjo, Buenos Aires, Argentina
17 June 1997 (Class 84/402); filed in Argentina 23 June 1995

This cymbal adds to the usual curved metal plate **2** a cup welded to the

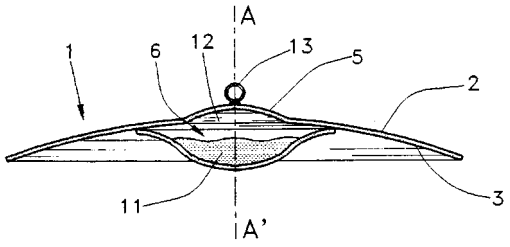


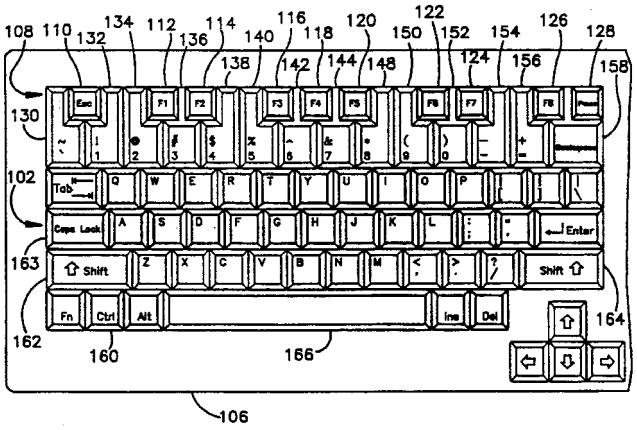
plate containing a liquid **11** and a gas **12**, in order to produce a different kind of tonal effect when the cymbal is struck.—DWM

5,646,648

43.75.Wx MUSICALLY ENHANCED COMPUTER KEYBOARD AND METHOD FOR ENTERING MUSICAL AND TEXTUAL INFORMATION INTO COMPUTER SYSTEMS

Randal Lee Bertram, assignor to International Business Machines Corporation
8 July 1997 (Class 345/168); filed 5 December 1994

This computer keyboard is enhanced by the addition of an upper row of keys resembling a piano style musical keyboard of approximately one and



one-half octaves. Function keys are also provided to enter musical information into the computer.—DWM

5,637,820

43.75.Yy STRINGED INSTRUMENT WITH ON-BOARD TUNER

Kenneth L. Wittman, Williamsport, PA
10 June 1997 (Class 84/454); filed 6 January 1995

This patent describes ''a stringed instrument with an on-board tuner which detects vibrations of the strings and determines and displays the identity of the string which is being played and the deviation of the string's pitch relative to an in-tune reference pitch. The tuner display is situated such that it is not readily visible to casual observers, such as an audience, yet is oriented such that the musician can easily view the tuner display from a normal playing position. In the case of an electric stringed instrument, the display is inlaid in the top surface of the neck of the instrument near the instrument's body. In the case of an acoustic stringed instrument, the tuner is mounted inside the body of the instrument and is visible through the sound hole.''—DWM

5,603,323

43.80.Qf MEDICAL ULTRASONIC DIAGNOSTIC SYSTEM WITH UPGRADEABLE TRANSDUCER PROBES AND OTHER FEATURES

Lauren S. Pflugrath and Jacques Souquet, assignors to Advanced Technology Laboratories
18 February 1997 (Class 128/660.01); filed 27 February 1996

This system is upgradeable with a new transducer hardware sent by air and by remote transmission of data that are used to control the operation of the transducer. The system includes communication hardware for reception of the upgrade data. For the upgrade, a communication link between the ultrasound system and a remote terminal is established using a common carrier.—RCW

5,615,679

43.80.Qf METHOD OF DISPLAYING ULTRASONIC IMAGES AND APPARATUS FOR ULTRASONIC DIAGNOSIS

Taiho Ri *et al.*, assignors to GE Yokogawa Medical Systems
1 April 1997 (Class 128/660.05); filed in Japan 6 February 1995

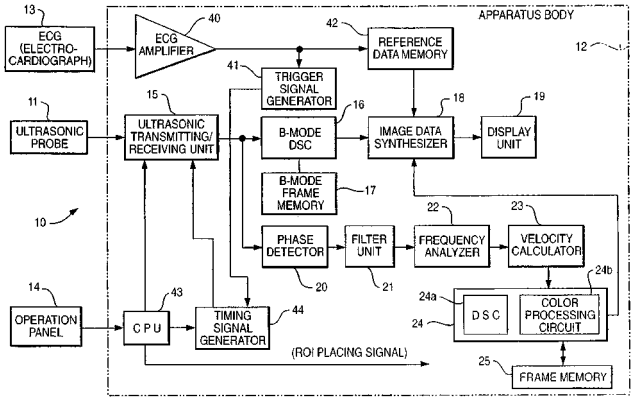
In this method and apparatus, an autocorrelator produces Doppler power data; a portion of the Doppler power image profile is extracted; the extracted portion of the image is shaded into values different from those of the remaining portion of the image; and the shaded image is visualized in three dimensions to facilitate diagnosis.—RCW

5,622,174

43.80.Qf ULTRASONIC DIAGNOSIS APPARATUS AND IMAGE DISPLAYING SYSTEM

Nobuo Yamazaki, assignor to Kabushiki Kaisha Toshiba
22 April 1997 (Class 128/661.09); filed in Japan 2 October 1992

In this apparatus and image display system patent (60 claims and 95 figures), movement information such as velocity and acceleration is ob-



tained about a tissue such as a cardiac muscle or the wall of a blood vessel in real time and displayed using color.—RCW

5,622,176

43.80.Qf METHOD FOR DETERMINING THE RISK OF TRISOMY 21 IN THE SECOND TRIMESTER

Anthony M. Vintzileos and James F. X. Egan, assignors to University of Medicine & Dentistry of NJ
22 April 1997 (Class 128/662.05); filed 26 July 1995

Trisomy 21 is a condition characterized by a small, anteroposteriorally flattened skull, short, flat-bridged nose, epicanthal fold, short fingers, and widened space between the first and second digits of hands and feet and

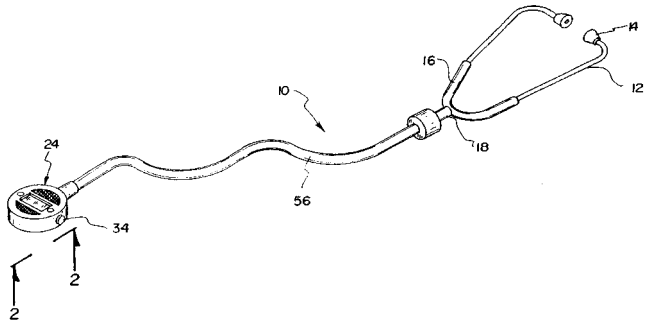
associated with moderate to severe retardation and a chromosomal abnormality. This method to detect trisomy 21 is based on fetal long boned biometry. Steps in the method are: ultrasonic measurement of the biparietal diameter and length of the femur, humorous, tibia, and fibula in fetuses of a patient population in the second trimester; amniocentesis on this population to determine which fetuses are normal and which fetuses have trisomy 21; derivation of equations that describe the predicted length of the measured bones based on biparietal diameter measurements, calculation of ratios of observed length to predicted lengths, comparison of the ratios for normal fetuses and for fetuses having trisomy 21; determination of a threshold for abnormally short bone length; and the use of the threshold to determine sensitivity and specificity for prenatal detection of trisomy 21.—RCW

5,638,453

43.80.Qf TRANSDUCER ENHANCED STETHOSCOPE

Bruce E. McLaughlin, Vista, CA
10 June 1997 (Class 381/67); filed 29 January 1996

The chestpiece on this battery operated electronic device has two faces with an LED display on each. One face shows the patient's skin temperature; the other shows his pulse rate. A reset button is provided to reset the pulse rate sensor. The chestpiece provides acoustical pickup of chest sounds.



These are delivered through a tube to an amplification unit near the binaurals. Very little detail on the acoustical and electronic elements is provided in the specification.—SFL

5,611,807

43.80.Sh ULTRASONIC ANGIOPLASTY BALLOON CATHETER

Matthew O'Boyle, Webster, TX
18 March 1997 (Class 606/169); filed 7 June 1995

The balloon in this catheter is piezoelectric and may be excited by application of an ultrasonic frequency signal. Balloon vibration caused by the excitation facilitates insertion of the balloon into a stenosis and also causes ablation of the stenosis. Inflation of the balloon in the stenosis does not predominantly cause a crushing of the material comprising the stenosis, but rather maintains contact between the stenosis and the vibrating surface to maximize delivery of ultrasonic energy for ablation.—RCW

5,601,082

43.80.Vj MEDICAL ULTRASOUND IMAGING

Christopher J. Barlow *et al.*, assignors to Intravascular Research Limited
11 February 1997 (Class 128/660.07); filed in the United Kingdom 30 September 1994

The ring down of an ultrasonic transducer is reduced by producing a reference scan, updating the scan on the basis of a long-term running average, and then subtracting the running average from the current a-scan.—RCW

5,601,083

43.80.Vj REAL TIME 3D IMAGING DEVICE USING FILTERED ELLIPSOIDAL BACKPROJECTION

Forrest Anderson, Bernalillo, NM
11 February 1997 (Class 128/660.07); filed 9 August 1993

This device includes an array of receiver elements each related to an image reconstruction point-angle of a receiver apodizer that weights echoes as a function of point-angle and an ellipsoidal backpropagation processor that reconstructs images from the weighted echoes.—RCW

5,601,085

43.80.Vj ULTRASOUND IMAGING

Jonny Ostensen *et al.*, assignors to Nycomed Imaging AS
11 February 1997 (Class 128/662.02); filed 2 October 1995

Successive images derived from echoes obtained in the presence of a contrast agent are correlated to determine areas where the correlation is poor, and the low level of correlation that results from contrast agent motion is used to distinguish between the contrast agent and relatively stationary bulk tissue since the echoes from bulk tissue are more highly correlated.—RCW

5,601,086

43.80.Vj BEAT FREQUENCY ULTRASONIC MICROSPHERE CONTRAST AGENT DETECTION SYSTEM

Robert A. Pretlow III *et al.*, assignors to the United States of America
11 February 1997 (Class 128/662.02); filed 12 May 1995

This system detects an ultrasonically reflective microsphere contrast agent by using the nonlinear sum and difference beat frequencies produced by the microspheres when two overlapping beams of signals with different frequencies are mixed by the nonlinear response of the microspheres.—RCW

5,603,324

43.80.Vj DUPLEXER INCLUDING A FIELD-EFFECT TRANSISTOR FOR USE IN AN ULTRASOUND IMAGING SYSTEM

Ralph Oppelt and Markus Vester, assignors to Siemens Aktiengesellschaft
18 February 1997 (Class 128/660.01); filed in Germany 19 May 1994

This duplexer couples a transducer element to either transmitting or receiving circuitry depending on whether the imaging system is in the transmit mode or receive mode. The gate of the field-effect transducer in this duplexer may either be driven with a control voltage or short-circuited to the source. Advantages of the duplexer are operation with low-power loss, without drive circuits, without leakage currents, and with low noise.—RCW

5,605,154

43.80.Vj TWO-DIMENSIONAL PHASE CORRECTION USING A DEFORMABLE ULTRASONIC TRANSDUCER ARRAY

Loriann L. Ries *et al.*, assignors to Duke University
25 February 1997 (Class 128/660.08); filed 6 June 1995

Correction of phase using this array is accomplished mechanically in one dimension and electronically in the other dimension, with the result that fewer electronic channels are needed for phase correction.—RCW

5,622,172

43.80.Vj ACOUSTIC DISPLAY SYSTEM AND METHOD FOR ULTRASONIC IMAGING

Ming Li and Jin Kim, assignors to Siemens Medical Systems, Incorporated
22 April 1997 (Class 128/661.1); filed 29 September 1995

Beamformed scan line signals in an ultrasonic imaging system with a three-dimensional display are separated into channel signals that are processed using head-related transfer function (HRTF) filters. An HRTF is associated with each of several points in an interrogated region. For each channel signal, a set of HRTF filter parameters is selected according to the sample volume from which the signal arises. Each HRTF parameter set defines left-ear and right-ear response functions to a calibration sound generated at a point that corresponds to a selected position in the interrogation region. Left and right audio output signals are produced to drive corresponding loudspeakers to which the user listens while performing an ultrasonic examination.—RCW

5,622,173

43.80.Vj COLOR FLOW DISPLAY WITH COMPENSATION FOR FLOW DIRECTION

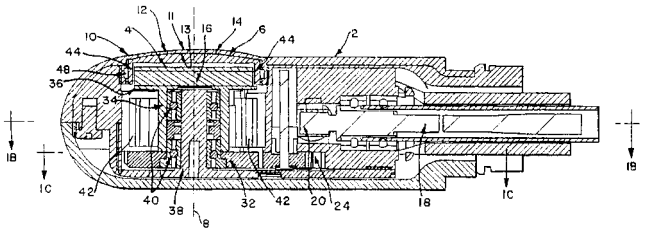
Janice Bisson and Lois Scheffler, assignors to Hewlett-Packard Company
22 April 1997 (Class 128/661.01); filed 21 December 1995

For this display, a user-controlled signal steers the angle of the ultrasonic beam. An image is produced using one color for flow in a direction toward the transducer, and an image using another color for flow away from the transducer. A processor responding to a steering signal that directs the ultrasonic beam from one side to the other side of an intermediate azimuth position produces a display of flow toward the transducer using the second color, and produces a display of flow in a direction away from the transducer using the first color when the beam is on the other side of the intermediate position. In this way, flow is represented with the same color when the beam-flow angle changes the apparent flow direction relative to the transducer. The colors are reversed when the flow changes direction.—RCW

5,626,138

43.80.Vj ULTRASONIC TRANSDUCER PROBE WITH AXISYMMETRIC LENS

John A. Hossack *et al.*, assignors to Acuson Corporation
6 May 1997 (Class 128/662.06); filed 7 June 1995



This probe contains an array of ultrasonic transducers **4** that can be rotated about an axis. A lens **6** is mounted on the array and a housing **2** encases the array and the lens. The lens is axisymmetric about the axis of rotation of the array and the housing has a continuous upper surface **10** that is formed to meet with the symmetric lens.—RCW

5,628,321

43.80.Vj PROCESSING VELOCITY INFORMATION IN AN ULTRASONIC SYSTEM

John P. Scheib and Sheng-Tz Lin, assignors to Diasonics Ultrasound, Incorporated
13 May 1997 (Class 128/661.08); filed 18 December 1995

Peak velocities are determined from spectra of echoes over a period of time. A time series of the velocities is determined to represent a so-called optimum cardiac cycle. A threshold magnitude is found by adding a predetermined value to a minimum magnitude of velocity. This threshold is used to find a peak velocity by a search. Information obtained from the optimum cardiac cycle includes peak systole, end diastole, minimum diastolic deflection, a rise time, an acceleration index, a pulsatility index, or a resistivity index.—RCW

5,628,322

43.80.Vj METHOD OF ULTRASOUND IMAGING AND DIAGNOSTIC ULTRASOUND SYSTEM

Yoshitaka Mine, assignor to Kabushiki Kaisha Toshiba
13 May 1997 (Class 128/661.08); filed in Japan 15 May 1995

Contrast echo imaging is accomplished to show the movement of a contrast medium injected into a patient intravenously. The system includes a unit that extracts the Doppler shift of an echo component nonlinearly related to the transmitted signal and attributable to reflection from the injected contrast agent. The Doppler shift is converted into a velocity using a reference signal that is preferably outside the band of the transmitted ultrasonic frequency.—RCW

Jovian acoustic matched-field processing[a)]

Michael D. Collins and B. Edward McDonald
Naval Research Laboratory, Washington, DC 20375

W. A. Kuperman
Scripps Institution of Oceanography, La Jolla, California 92093

William L. Siegmann
Rensselaer Polytechnic Institute, Troy, New York 12180

(Received 17 August 1996; accepted for publication 31 March 1997)

Evidence of waves has been observed in Hubble Space Telescope (HST) images of the Comet Shoemaker–Levy 9 impact sites [H. B. Hammel *et al.*, ''HST imaging of atmospheric phenomena created by the impact of Comet Shoemaker–Levy 9,'' Science **267**, 1288–1296 (1995)]. The radius of a ring that appears near the fragment G impact site is consistent with an acoustic wave and increases with time. This feature cannot be explained in terms of a naive plane-wave model because it appears to expand at a rate that is slower than the speed associated with its radius. This behavior can be explained in terms of multipath propagation in the sound channel. According to this hypothesis, acoustic waves corresponding to different group speeds followed paths that passed in and out of the relatively thin debris layer in which the feature appears. Matched-field processing for the depth of the source indicates that the fragment G explosion must have occurred deep in the water clouds if the ring indeed corresponds to an acoustic wave. © *1997 Acoustical Society of America.* [S0001-4966(97)00910-7]

PACS numbers: 43.10.Ln, 43.60.Gk [JLK]

INTRODUCTION

Expanding rings near the impact sites of the fragments of Comet Shoemaker–Levy 9 appear in Hubble Space Telescope (HST) images of Jupiter. A small ring that appears near several impact sites expanded at the constant rate of about 450 m/s for more than an hour.[1] Although this behavior is consistent with a wave, it has not been reproduced using a wave model and environmental profiles from the Voyager[2] and Galileo missions. The nature of the small ring has not been determined.

In this paper, we consider a large ring that is concentric to and about twice as large as the small ring. The large ring appears in a few of the methane-band images of the fragment G impact site. It has not been determined whether this expanding feature corresponds to moving debris[3] or a wave. However, the moving debris hypothesis seems to be inconsistent with the fact that the debris from the fireball plume appeared to decelerate rapidly when it collapsed onto the stratosphere more than an hour before the rings were imaged.[1]

Although the radius of the large ring is consistent with an acoustic wave propagating in the sound channel,[4] the rate of expansion appears (within the naive framework of a plane wave) to be subsonic during a 34-min interval between images. We show that this behavior is consistent with acoustic waves corresponding to different group speeds that follow paths that pass in and out of the relatively thin debris layer[5] from above and below. Under this hypothesis, the waves produce visible effects when they pass through the debris layer. We perform matched-field processing[6,7] for the depth of the source and show that fragment G must have exploded deep in the water clouds if the large ring indeed corresponds to acoustic waves.

I. OUTWARD MOTION

HST images of the fragment G impact site appear in Fig. 1. The significant differences between the images are largely due to the fact that the green image includes features at relatively low altitudes. The features of interest appear dark in the green image and bright in the methane-band image and include the fragment G impact spot near the center, the asymmetric debris field to the lower right, the fragment D impact spot to the left of the center, the small ring, and the large ring. The rings that are centered about the fragment G impact spot are circular but appear elliptical due to the viewing angle. The small ring is prominent in both images. It expands at about 450 m/s, has a radius that is slightly greater than half the distance to the fragment D spot, and is relatively weak in the direction opposite the debris field. The large ring is fairly prominent in the methane-band image but absent from (or very weak in) the green image. It appears only in the debris field and has a radius that is slightly larger than the distance to the fragment D spot. This radius is consistent with the group speeds of acoustic waves propagating in the sound channel.[4] The color image appearing in Fig. 1 of Ref. 5 contains a subtle circular feature that is of the same radius as the large ring and spans most of the region opposite the debris field.

Different analyses are required for the small and large rings due to differences in the available data for these fea-

[a)]''Selected research articles'' are ones chosen occasionally by the Editor-in-Chief that are judged (a) to have a subject of wide acoustical interest, and (b) to be written for understanding by broad acoustical readership.

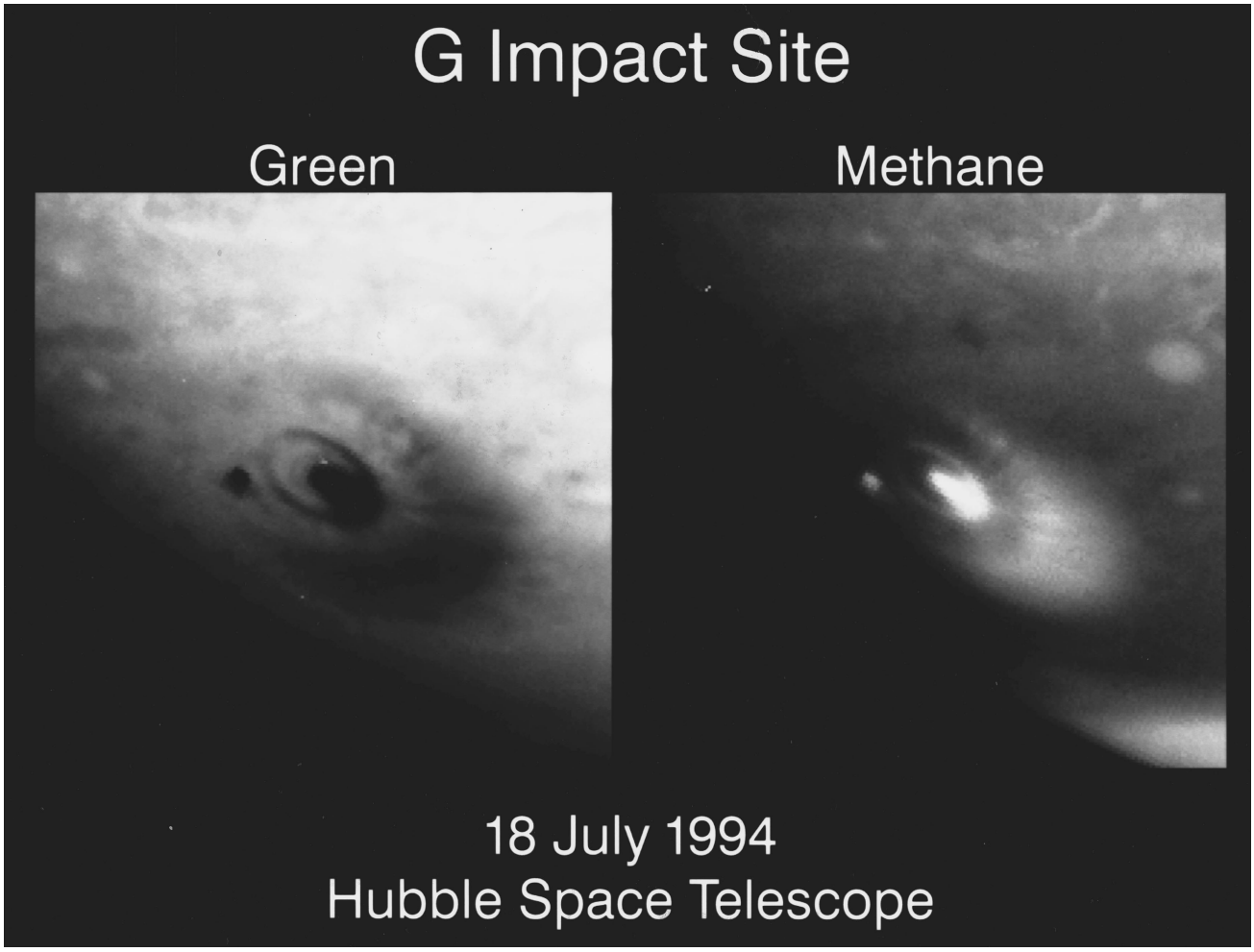


FIG. 1. HST images of the fragment G impact site. Features of interest appear dark in the green image and bright in the methane-band image. The fragment G impact site corresponds to the spot near the center. The fragment D impact site corresponds to the spot to the left of the fragment G impact site. The debris field is to the lower right. The rings appear elliptical due to the viewing geometry. The small ring is prominent in both images and has a radius that is slightly more than half the distance between the fragments G and D impact sites. The large ring is fairly prominent in the methane-band image, absent from (or very weak in) the green image, and concentric to and about twice as large as the small ring. [Images courtesy H. B. Hammel and NASA.]

tures and differences in the behavior of acoustic and gravity waves. The small ring appears at several wavelengths, in several images, and near several impact sites. Having a significant number of data points, Hammel *et al.* were able to reliably estimate the speed of the small ring by least squares and show that the speed is nearly independent of the distance from the impact site and the strength of the source.[1] This suggests that the small ring corresponds to a linear wave. Since the observed speed is well below the minimum group speed of acoustic waves, Hammel *et al.* considered the possibility that the small ring corresponds to a gravity wave. Further analysis was required to explain this feature, however, because the sharpness of the observed wave front is not consistent with the wide range of horizontal group speeds corresponding to the profiles constructed from Voyager data.[2] Prior to arrival of Galileo at Jupiter, Ingersoll and Kanamori proposed an enhanced water cloud that supports a gravity wave dominated by a single mode.[8] Since the Galileo data do not support this hypothesis, the nature of the small ring remains a mystery. Ongoing efforts to explain the small ring in terms of an acousto-gravity wave seem promising because 450 m/s is relatively large for a pure gravity wave (incompressibility is equivalent to the acoustic wave speed being much greater than the gravity wave speed).

A least-squares analysis is not appropriate for the large ring because it has been observed in only a few images of the fragment G impact site. If the large ring corresponds to a wave, however, its speed can be estimated by building on the results of Hammel *et al.* Since the most prominent ring behaves like a linear wave away from the impact sites, it is reasonable to assume that all waves behave linearly away from the impact sites. Under this assumption, the speed of a wave feature can be estimated from a single observation of its radius at a given time after impact. A small correction can be included by accounting for nonlinear effects near the source by using a time shift or an effective source radius, which is about 600 km for the small ring.[1] For the large ring, this approach gives a range of speeds that is consistent with the group speeds of about 800–1200 m/s corresponding to acoustic waves propagating in the sound channel. This agreement holds for the unmodified Voyager profiles. It was necessary to modify the Voyager profiles to match a gravity wave to the small ring.[8]

While searching for additional evidence that the large

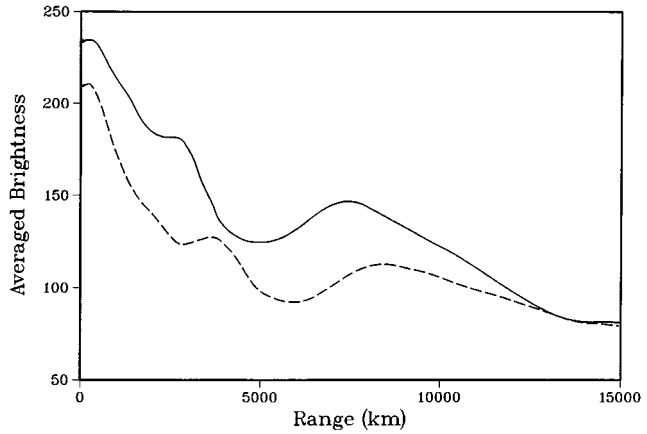


FIG. 2. Plots of brightness versus range averaged over azimuth from northeast to southwest in HST methane-band images that were obtained 91 min (solid curve) and 125 min (dashed curve) after the impact of fragment G. The peak that moves from $r=2970$ km to $r=3900$ km corresponds to the small ring. The peak that moves from about $r=7400$ km to $r=8400$ km in range corresponds to the large ring, which is consistent with acoustic waves. [Data courtesy of T. E. Dowling.]

ring corresponds to an acoustic wave, we came across evidence that at first seemed to contradict this hypothesis. We measured the radius of the large ring in methane-band images that were taken 91 and 125 min after the impact of fragment G (Fig. 1 of Ref. 1). The apparent radius of the large ring increased from about 7400 to 8400 km during the 34-min interval between images. This motion is illustrated in the plots of brightness versus range that appear in Fig. 2. The prominent peaks in brightness that occur at the ranges of the small and large rings move outward during the interval between the images. Although the expansion of the large ring is consistent with a wave, the rate of expansion indicates a speed of about 490 m/s, which is far below the minimum group speed of acoustic waves. This would seem to eliminate wave motion as an explanation for the expansion of the large ring.

Besides wave motion, the only reasonable explanation for the expansion of the large ring would be lateral motion of debris,[3] which has been observed in HST images[1] and apparently accounts for the shape of the debris field.[9] Debris motion has also been proposed as an explanation for an expanding ring located beyond the debris field.[10] For the moving debris hypothesis to work, the large ring had to be located at the same position that an acoustic wave would be expected and the speed of the material had to be on the order of 500 m/s more than 90 min after impact. A coincidental location would complicate the interpretation of the large ring but would not be surprising. The other requirement is harder to accept because the lateral motion of debris appeared to decelerate rapidly and stop just after the debris fell back onto the stratosphere about 20–30 min after impact in sequences of HST images (Fig. 2 and Table 3 of Ref. 1).

II. ESTIMATION OF THE IMPACT DEPTH

Since the moving debris hypothesis has not been established, we continued to investigate the acoustic wave hypothesis and found that the multipath nature of acoustic propaga-

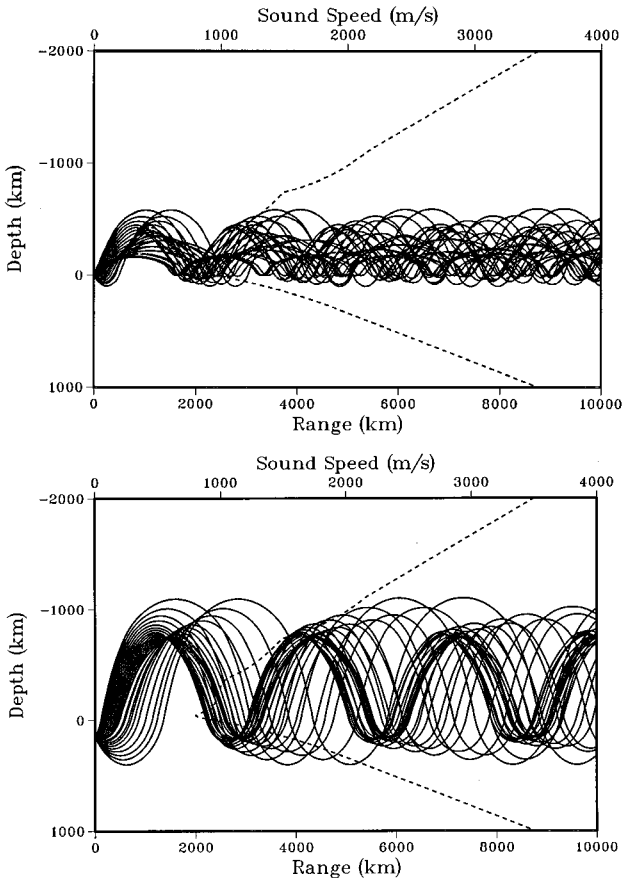


FIG. 3. Ray paths in the Jovian sound channel for sources at 1 bar (top) and 25 bar (bottom). Depth is the distance below the 1-bar level (which is near the tropopause). Rays launched within 45 deg of horizontal are included. The rays for the 1-bar source correspond to average horizontal speeds of 885–970 m/s and are confined to a relatively thin layer in the sound channel. The rays for the 25-bar source correspond to average horizontal speeds of 960–1270 m/s. The sound-speed profile that appears as a dashed curve was generated by linearly extrapolating and smoothing a profile obtained from Voyager data.[2]

tion in the sound channel might be the key to understanding the outward motion. The ray paths in Fig. 3 illustrate how acoustic energy repeatedly oscillates about the tropopause while it propagates out in range. As the rays cycle about the tropopause, they spend relatively little time within the debris field, which occupies a relatively thin layer in the sound channel between about 2 and 200 mbar (Ref. 5) and where waves might produce visible effects. This is especially true for the source far below the tropopause at 25 bar. At a given observation point, the acoustic field due to a pulsed source consists of a sequence of arrivals. The earliest arrivals correspond to rays that radiate from the source at relatively steep angles and propagate far above and below the tropopause, where the speed of sound is relatively high. The latest arrivals correspond to energy trapped near the tropopause. Similar behavior has been modeled and observed in the ocean sound channel.[11]

The arrival times and amplitudes depend on the location of the source and the receiver (or observation point). For an acoustic wave observed within the debris field, all of these position parameters would be known except for the vertical location of the source. We have qualitatively estimated the

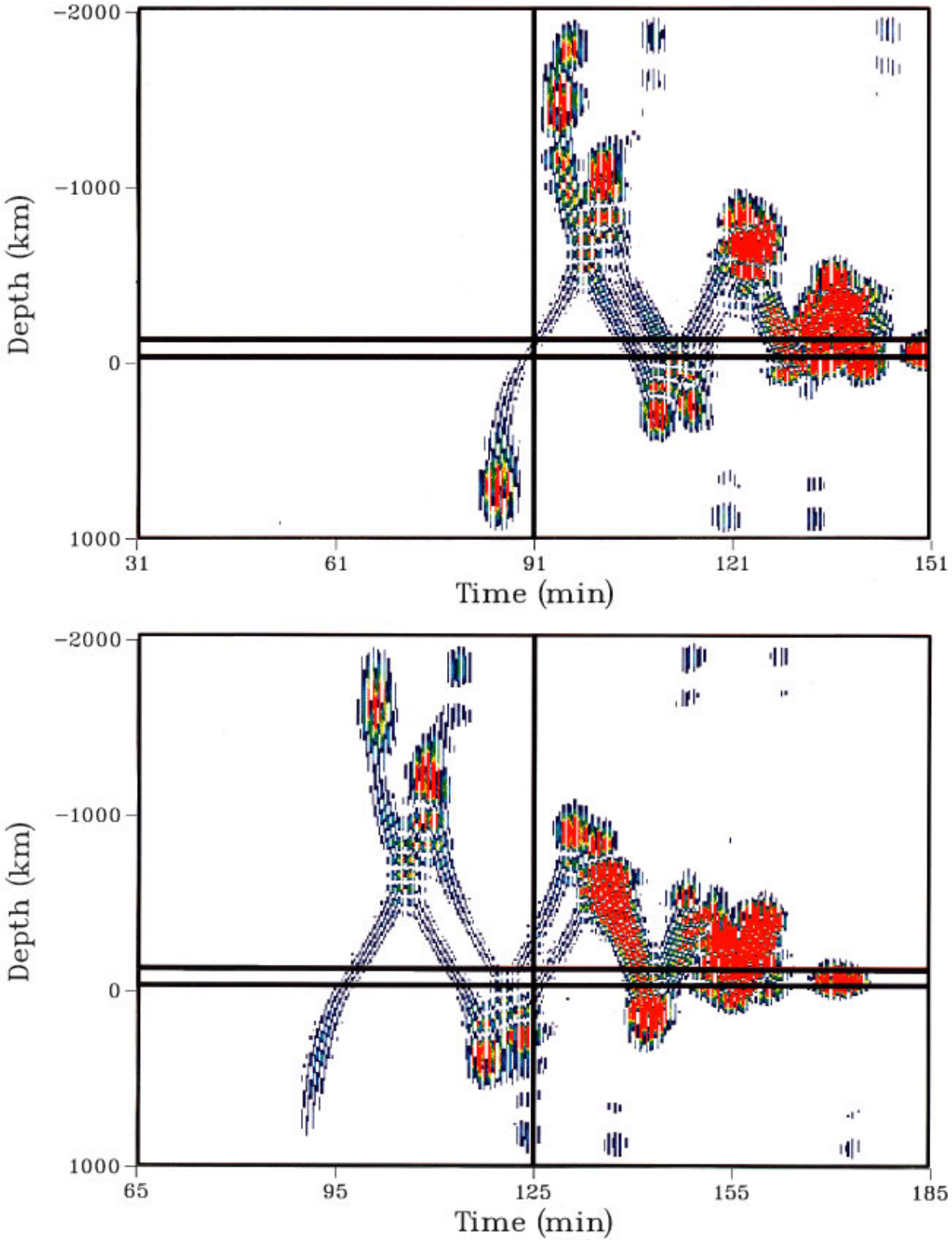


FIG. 4. Modeled acoustic time series corresponding to a source at 1 bar that are received at $r = 7400$ km (top) and $r = 8400$ km (bottom). Red corresponds to high intensity. Blue corresponds to low intensity. The debris layer lies between the horizontal lines. The vertical lines denote 91 min (top) and 125 min (bottom) after the pulse is emitted. HST images that contain the large ring were obtained at these times.

vertical location of the fragment G explosion that is consistent with the acoustic wave hypothesis. In ocean acoustics, this inverse problem (which in general involves all three source coordinates) is known as matched-field processing.[6,7] Source localization in free space is relatively simple because there is only one arrival, which can be used to determine the direction of the source. Source localization in a waveguide is complicated by multipath propagation. This complexity can be an advantage, however, because each arrival contains information about the location of the source. In fact, it is pos-

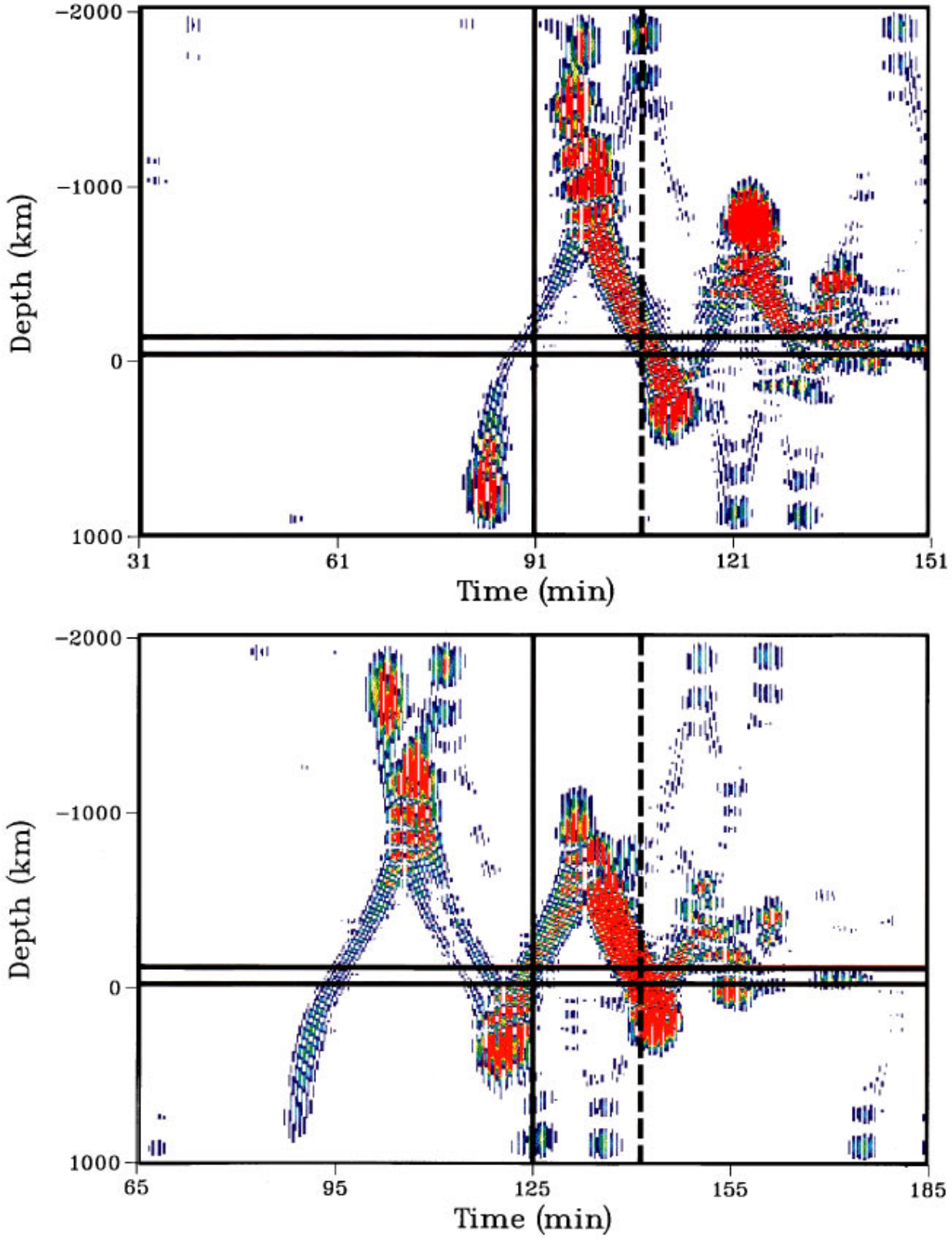


FIG. 5. Modeled acoustic time series corresponding to a source at 25 bar that are received at $r = 7400$ km (top) and $r = 8400$ (bottom). Red corresponds to high intensity. Blue corresponds to low intensity. The debris layer lies between the horizontal lines. The solid vertical lines denote 91 min (top) and 125 min (bottom) after the pulse is emitted. HST images that contain the large ring were obtained at these times. As the dashed vertical lines indicate, the largest arrivals in the debris layer occur at both locations about 16 min after the HST images were obtained. This time shift could be due to nonlinear behavior near the source.

sible to localize a source in a waveguide with a single receiver.[12] It is necessary to use an array of receivers to localize a source in free space.

We work in cylindrical coordinates, where the range r is the horizontal distance from the source and z is the depth below the 1-bar level. We neglect the curvature of the waveguide, the zonal winds,[13] and coupling of energy between planes of constant azimuth because these factors are not important at short ranges. We neglect refraction due to the ambient density gradient because it is relatively weak at the frequencies of interest.[4] We neglect nonlinear effects, which are important near the source and at high altitudes. We solve the forward problem by Fourier synthesizing frequency-domain solutions obtained using the parabolic equation method.[14] We designed the rational approximation associated with the parabolic equation method to be accurate for energy trapped in the sound channel and to annihilate steeply propagating energy that reflects from the boundaries of the computational domain at $z=-2000$ km and $z=1000$ km. We generated an initial condition at $r=0$ using Greene's starting field,[15] which should be sufficient for qualitative purposes. The sound-speed profile obtained from Voyager data[2] only covers -1109 km$<z<273$ km. The sound-speed profile appearing in Fig. 3 was obtained over the computational domain by linearly extrapolating the Voyager profile and smoothing. The extrapolated portions of the profile do not have a significant effect on the acoustic modes of interest.

Although the source signature is not known, a generic pulsed source can be used to investigate the key elements of the localization problem. We consider acoustic frequencies slightly higher than the Brunt–Väisälä frequency[16] and use the source function $\cos(2\pi ft)\exp[-(t/\tau)^2]$, where $f=10$ mHz and $\tau=150$ s. We computed acoustic time series at $r=7400$ km and $r=8400$ km and searched for the source depth that is most consistent with the observations. The time series appearing in Figs. 4 and 5 correspond to sources at 1 and 25 bar. The arrival structure in these time series can be interpreted in terms of the ray paths appearing in Fig. 3. Different arrivals correspond to rays that cycle above and below the tropopause different numbers of times. Similar times series have been modeled and observed in the ocean sound channel.[11] The arrival structure corresponding to the 1-bar source is not consistent with the observations. For both ranges, most of the energy is concentrated into a relatively thin layer of the sound channel (as the ray paths in Fig. 3 suggest), and the strongest arrival in the debris field occurs long after the corresponding image was obtained. When shifted by 16 min, the arrival structure corresponding to the 25-bar source is consistent with the observations. For both ranges, the strongest arrival in the debris field occurs about 16 min after the corresponding HST image was obtained. The strongest arrival at $r=7400$ km corresponds to relatively steep rays and high group speeds. The strongest arrival at $r=8400$ km corresponds to rays trapped near the tropopause. A 16-min shift due to nonlinear behavior near the source would be reasonable because the effective size of the 450-m/s wave source is estimated to be 586 ± 125 km (Ref. 1), which corresponds to a time shift of 21.7 ± 4.6 min. It is necessary to place the source between about 20 and 30 bar in order for the strongest arrivals to line up as they do in Fig. 5. The strongest arrivals do not come close to lining up in Fig. 4.

The agreement with the observations is very good for such a simple inverse problem involving only one parameter (the source depth). The agreement could presumably be improved by including corrections to the sound-speed profile in the parameter search. This approach is effective for solving matched-field processing problems in ocean acoustics.[17] Other possible (but difficult) ways to improve the agreement would be to model nonlinear effects near the source and to allow the source to be directional. Although the inverse problem for the source depth is very simple, the forward problem of guided waves in the sound channel is fairly complicated. This is an advantage because false matches become less likely as the complexity of the forward problem increases.[18,19] False matches are unlikely for this problem because the acoustic field consists of a sequence of arrivals that vary in a complicated fashion as the source depth varies. Based on the simplicity of the localization problem and the complexity of the forward problem, it would be safe to conclude that fragment G exploded deep in the water clouds if the large ring is conclusively identified as an acoustic wave.

Determining the nature of the large ring will require further data and analysis. One approach would be to search for further evidence of the large ring that appears to the northwest of the fragment G impact site (Fig. 1 of Ref. 5). Ruling out the possibility that this feature consists of debris (this seems reasonable because most of the debris is to the southeast) would lead to the conclusion that the feature is an acoustic wave and that acoustic waves are detectable outside the debris field. If this turns out to be the case, the caustic that forms to the west of the impact site[4] might be detectable because enhancement in the caustic compensates for cylindrical spreading loss. The multipath nature of acoustic waves should also be considered for any moving features that are detected outside the large ring. Any such features that correspond to acoustic waves interacting with a relatively thin layer of the atmosphere will consist of multipaths and appear to move at a variable rate.

III. CONCLUSIONS

A large ring is visible within the debris field in methane-band images and possibly outside the debris field in a color image. During a 34-min interval between images, this feature appears to expand at about 490 m/s, which is well below the minimum group speed of acoustic waves. Advection is an alternative explanation for the motion of the large ring. However, sequences of HST images suggest that lateral motion stopped abruptly after the debris field was formed. The behavior of the large ring is consistent with acoustic waves associated with different horizontal group speeds passing in and out of the debris layer as they follow paths that oscillate vertically in the sound channel. We performed matched-field processing for the source depth and found that the motion of the large ring is consistent with acoustic waves that originated deep in the water clouds. To achieve agreement with the observations, it is necessary to assume a time shift of about 16 min due to nonlinear effects near the source. A

comparable shift (or equivalent finite source size) is required to explain the small ring in terms of a wave.[1]

ACKNOWLEDGMENTS

We thank G. S. Orton for providing the Voyager data and T. E. Dowling for providing the HST data. This work was supported by the Office of Naval Research.

[1] H. B. Hammel, R. F. Beebe, A. P. Ingersoll, G. S. Orton, J. R. Mills, A. A. Simon, P. Chodas, J. T. Clarke, E. De Jong, T. E. Dowling, J. Harrington, L. F. Huber, E. Karkoschka, C. M. Santori, A. Toigo, D. Yeomans, and R. A. West, ''HST imaging of atmospheric phenomena created by the impact of Comet Shoemaker–Levy 9,'' Science **267**, 1288–1296 (1995).

[2] G. F. Lindal, G. E. Wood, G. S. Levy, J. D. Anderson, D. N. Sweetnam, H. B. Hotz, B. J. Buckles, D. P. Holmes, P. E. Doms, V. R. Eshleman, G. L. Tyler, and T. A. Croft, ''The atomosphere of Jupiter: An analysis of the Voyager Radio Occultation Measurements,'' J. Geophys. Res. **86**, 8721–8727 (1981).

[3] K. Zahnle, ''Models of fallback,'' Presented at IAU Colloquium 156 (Baltimore, Maryland, 1995).

[4] M. D. Collins, B. E. McDonald, W. A. Kuperman, and W. L. Siegmann, ''Jovian acoustics and Comet Shoemaker–Levy 9,'' J. Acoust. Soc. Am. **97**, 2147–2158 (1995).

[5] R. A. West, E. Karkoschka, A. J. Friedson, M. Seymour, K. H. Baines, and H. B. Hammel, ''Impact debris particles in Jupiter's stratosphere,'' Science **267**, 1296–1301 (1995).

[6] H. P. Bucker, ''Use of calculated sound fields and matched-field detection to locate sound sources in shallow water,'' J. Acoust. Soc. Am. **59**, 368–373 (1976).

[7] A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, ''An overview of matched field methods in ocean acoustics,'' IEEE J. Ocean Eng. **18**, 401–424 (1993).

[8] A. P. Ingersoll and H. Kanamori, ''Waves from the collisions of comet Shoemaker–Levy 9 with Jupiter,'' Nature **374**, 706–708 (1995).

[9] E. M. Shoemaker, P. J. Hassig, and D. J. Roddy, ''Numerical simulations of the Shoemaker–Levy 9 impact plumes and clouds: A progress report, '' Geophys. Res. Lett. **22**, 1825–1828 (1995).

[10] P. J. McGregor, P. D. Nicholson, and M. G. Allen, ''CASPIR observations of the collision of Comet Shoemaker–Levy 9 with Jupiter,'' Icarus **121**, 361–388 (1996).

[11] T. F. Duda, S. M. Flatté, J. A. Colosi, B. D. Cornuelle, J. A. Hildebrand, W. S. Hodgkiss, P. F. Worcester, B. M. Howe, J. A. Mercer, and R. C. Spindel, ''Measured wave-front fluctuations in 1000-km pulse propagation in the Pacific Ocean,'' J. Acoust. Soc. Am. **92**, 939–955 (1992).

[12] L. N. Frazer and P. I. Pecholcs, ''Single-hydrophone localization,'' J. Acoust. Soc. Am. **88**, 995–1002 (1990).

[13] S. S. Limaye, ''Jupiter: New estimates of the mean zonal flow at the cloud level,'' Icarus **65**, 335–352 (1986).

[14] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, Woodbury, NY, 1994), pp. 343–412.

[15] R. R. Greene, ''The rational approximation to the acoustic wave equation with bottom interaction,'' J. Acoust. Soc. Am. **76**, 1764–1773 (1984).

[16] A. E. Gill, *Atmosphere-Ocean Dynamics* (Academic, San Diego, 1982), p. 129.

[17] M. D. Collins and W. A. Kuperman, ''Focalization: Environmental focusing and source localization,'' J. Acoust. Soc. Am. **90**, 1410–1422 (1991).

[18] J. S. Perkins and W. A. Kuperman, ''Environmental signal processing: Three-dimensional matched-field processing with a vertical array,'' J. Acoust. Soc. Am. **87**, 1553–1556 (1990).

[19] M. D. Collins, L. T. Fialkowski, W. A. Kuperman, and J. S. Perkins, ''Environmental source tracking,'' J. Acoust. Soc. Am. **94**, 3335–3341 (1993).

Acoustic wave propagation through porous media: Theory and experiments[a)]

Tim W. Geerits[b)] and Oscar Kelder
Faculty of Applied Earth Sciences, Delft University of Technology, P.O. Box 5028, 2600 GA Delft, The Netherlands

(Received 5 May 1994; revised 17 January 1997; accepted 14 May 1997)

A new lossless poroelastic wave propagation theory is verified by means of ultrasonic transmission measurements on artificial rock samples in a water-filled tank. The experiments involved are similar to those performed by Plona [Appl. Phys. Lett. **36**, 259–261 (1980)]. In this physically and mathematically mutual consistent new theory the coupling terms between the fluid and solid phase of the porous medium are completely determined by the measured wave speeds and the mass densities and constitutive parameters of both constituting phases. Verification of the amplitudes of the received bulk waves in both the time domain and frequency domain provide information on the validity of the combined effect of propagation characteristics and new macroscopic fluid/fluid-saturated-rock boundary conditions resulting from this theory. The comparison technique between theory and experiments is based on the Fraunhofer diffraction theory, and is first tested in a perfectly elastic medium transmission configuration. Subsequently, this comparison technique is used for the poroelastic medium. It is shown that this technique is very accurate and reliable. The experimental results for the compressional wave and the shear wave in the perfectly elastic medium are in excellent agreement with the theoretical predictions. For the fluid-saturated porous samples, good agreement is found only for the fast compressional wave. For both the shear wave and the slow compressional, it is obvious that there is some kind of loss mechanism involved, which cannot be explained by the current theory. Despite the fact that the bulk losses in the porous medium can be explained qualitatively by the full frequency range Biot theory, it is conjectured that even a quantitative fit is feasible if Johnson's loss model [D. L. Johnson *et al.*, J. Fluid Mech. **176**, 379–402 (1987)] is applied in the lossy counterpart of the current theory [T. W. Geerits, J. Acoust. Soc. Am. **100**, 2949–2959 (1996)].
[S0001-4966(97)00610-3]

PACS numbers: 43.10.Ln, 43.20.Gp, 43.20.Jr, 43.40.Ph [JEG]

INTRODUCTION

In the exploration and exploitation of the Earth's natural resources, in particular oil and gas, it has been attempted to determine rock properties from acoustic measurements on the subsurface. The inverse of this process, often referred to as ''the forward model,'' predicts the measurements if the rock parameters (medium parameters) and the source and receiver characteristics are known. At the basis of the forward model there is some kind of theory which describes the propagation of transient acoustic waves through a medium, representative of the subsurface. In this context such a medium may be considered a fluid-saturated porous medium. The aim of this paper is to verify a new lossless poroelastic wave propagation theory for such a fluid-saturated porous medium. This lossless theory was first presented by de Vries[1] and extended by Geerits for the case in which losses of the interface type are incorporated.[2,3] The resulting field equations of this lossless theory have the same appearance as the high-frequency limit Biot equations[4–6] but a different physical interpretation. As predicted by Biot in 1956, a fluid-saturated porous medium is characterized by the propagation of both a fast and a slow compressional wave, as well as a shear wave. The first clear observation of a propagating slow compressional wave was reported by Plona[7] for water-saturated sintered glass beads. Subsequently, more extensive quantitative experimental research on these kind of porous materials show the strong predictive power of the Biot theory, and confirms the bulk attenuation mechanism involved.[8–11] Similar to the Biot theory, this new poroelastic wave propagation theory also predicts the existence of these three bulk waves. However, there are two main differences between this theory and the classical phenomenological Biot theory, later derived from microstructure by Burridge and Keller[12] and Pride *et al.*[13] First, the present theory is physically and mathematically mutual consistent thereby having its own physical interpretation and therefore deviant macroscopic boundary conditions (fluid/porous, solid/porous, porous/porous) in comparison with the more classical ones proposed by Deresiewicz and Skalak[14] for the Biot theory, and second, its derivation from microstructure by means of spatial volume averaging which is much easier to comprehend than the method of space homogenization used by

[a)]''Selected research articles'' are ones chosen occasionally by the Editor-in-Chief that are judged (a) to have a subject of wide acoustical interest, and (b) to be written for understanding by broad acoustical readership.
[b)]Present address: Western Atlas International, 10201 Westheimer, Houston, TX 77042.

Burridge and Keller.[12] An explanation of these differences was recently published by Geerits.[3] As mentioned before, verification of the acoustic properties (reflection/transmission and bulk attenuation/speed) of fluid-saturated porous materials is not new anymore. However, the original and interesting observations of our work are, first, the use of a new poroelastic theory in which the coupling between the fluid and solid phase of the porous medium is completely determined by the measurable wave speeds of both the porous medium (FP, SP, and S) and its constituents, the mass densities, and the constitutive parameters of both constituting phases, whereas the Biot theory introduces several rock and fluid properties to describe this fluid/solid coupling. Second, the verification of the combined effect of propagation characteristics and new macroscopic boundary conditions on all three measured bulk waves, and, third, the high accurate and very reliable comparison technique based on the Fraunhofer diffraction theory. To verify both the propagation characteristics of the bulk waves and the new macroscopic boundary conditions resulting from the theory, a transmission configuration is used, similar to the one used by Plona. A special case of this poroelastic theory is the perfectly elastic solid theory.[15] To show the accuracy of the measurement configuration, this perfectly elastic solid theory will be verified first.

Section I describes the transmission configuration which we use for the verification of the bulk waves in both the perfectly elastic solid and the fluid-saturated porous medium. Section II gives the basic relations which describe the acoustic wave propagation through the fluid and, respectively, the perfectly elastic solid and the porous medium.

In Sec. III the relevant field equations will be subjected to two integral transformations: one with respect to time (*Laplace* transformation) and one with respect to the horizontal space coordinates (*Fourier* transformation). These transformations take advantage of the fact that the configuration is time invariant and shift invariant in the two horizontal directions. In this so-called transformed domain the propagation of the acoustic waves is brought back to one spatial dimension of which the direction is perpendicular to the fluid/solid (porous) interface.[16,17] To understand the mathematical structure of the relevant field equations in the transformed domain, the transformed field equations are put in a matrix formalism,[1] according to the well-known reflectivity method of Kennett.[18] After diagonalization of the system matrix it is possible to write the transformed field quantities as a weighted sum of the different bulk wave contributions, in which both the transmission and reflection coefficients of the different bulk waves at the fluid/perfectly-elastic-solid, and fluid/fluid-saturated-porous-medium interfaces can be identified.

In Secs. IV and V, the macroscopic boundary conditions at both the fluid/perfectly-elastic-solid/fluid, and fluid/fluid-saturated-porous-medium/fluid interfaces will be discussed and compared with the classical boundary conditions for the Biot theory. From these boundary conditions, a new matrix formalism is formulated from which we derive the net transmission coefficient for the entire slab of elastic, porous medium. This transmission coefficient turns out to be a summation of the angular wave-number-dependent amplitude factors of the different bulk waves. Furthermore, this matrix formalism is solved such that it offers the opportunity to decide whether multiples (and how many) should be taken into account. In this paper multiples will be excluded, i.e., we only consider primaries. Once the transmission coefficient at the second interface (i.e., at the receiver side) has been calculated, the transformed acoustic traction (opposite of pressure) will be calculated at an arbitrary receiver position in the lower fluid half-space.

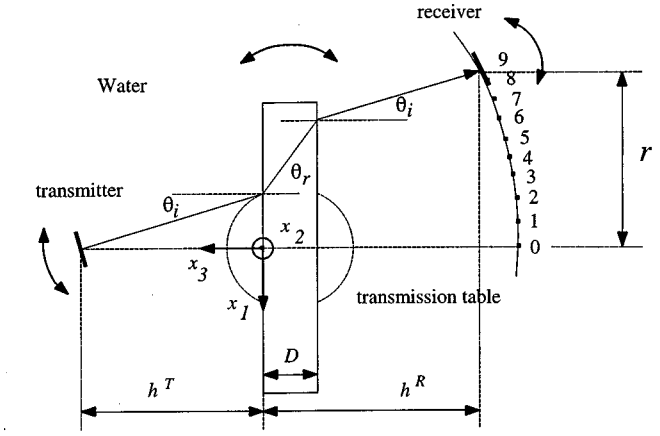


FIG. 1. The configuration.

Section VI summarizes the transformation of the transformed acoustic traction back to the space-time domain. This procedure will be executed, analytically, by means of the *Cagniard–de Hoop* method.[17,19,20] This integral transformation technique is exact and perfectly suited for this type of configuration.

Subsequently, Sec. VII provides a summary of the *Fraunhofer* diffraction theory[21] as far as it is relevant in the context of this paper. The pressure field emitted by a point source in a homogeneous and isotropic ideal (frictionless) fluid of infinite extent will be compared with the pressure field emitted by a flat, piezoelectric transducer in the far field. From this comparison the experimental configuration conditions are deduced under which the actual measured seismograms can be compared with the synthetic seismograms. Furthermore, the synthetic seismograms will be compared with the measurements and the results will be discussed. Finally, conclusions with suggestions for future research will be given in Sec. VIII.

I. THE CONFIGURATION

In Fig. 1 a top view of the configuration is presented, where we have used a Cartesian coordinate system such that the coordinates of the transmitter $\mathbf{x}^T$ are always given by $(0,0,h^T)$. The quantity h^T denotes the distance between the fluid/solid (porous) interface and the transmitter, perpendicular to the interface. The horizontal spatial coordinates coincide with the fluid/solid, porous interface plane. Both transmitter and receiver can be rotated in the x_1-x_3 plane. The transmitter is fixed at one position while the receiver is mounted in a x_1, x_2, x_3 positioning system. Between the

transmitter and receiver a slab of perfectly elastic medium (steel) or a slab of porous material (fused glass beads), of thickness D is situated, which can be rotated on a transmission table to any desired angle. All translations and rotations are fully computerized, and the entire setup is situated in a water tank. The transmission object divides the total fluid space in two half-spaces: an upper fluid half-space ($x_3>0$) and a lower fluid half-space ($x_3<-D$). Furthermore, some additional configuration coordinates are shown in Fig. 1. They will be frequently used throughout this paper.

II. BASIC RELATIONS

Both the upper and lower fluid half-space are assumed to be ideal (frictionless) while the slab of steel is assumed to be perfectly elastic. For the porous medium we use the homogeneous, isotropic porous medium equations as given by de Vries.[1–3] The relevant field equations may be summarized as follows.

A. Ideal (frictionless) fluid equations

$$\partial_i \sigma + f_i^f = \rho^f \, \partial_t v_i^f, \tag{1}$$

where σ is the omnidirectional fluid traction [Pa], ϱ^f is the omnidirectional fluid mass density [kg/m^3], f_i^f is the volume source density of fluid body force [N/m^3], and v_i^f is the fluid particle velocity [m/s]. Further

$$\kappa \, \partial_t \sigma = \partial_k v_k^f - h^f, \tag{2}$$

where κ is the isothermal compressibility [Pa^{-1}], and h^f is the fluid volume injection source rate [s^{-1}]. Equation (1) represents the linearized equation of motion which pertains to an isotropic and homogeneous ideal (frictionless) fluid, i.e., only small disturbances relative to an equilibrium situation are considered. Equation (2) follows from combining the continuity equation and the definition of isothermal compressibility.[15] The resulting equation (equation for isothermal flow) is then linearized. Equation (2) may be considered the constitutive relation for an isotropic and homogeneous ideal (frictionless) fluid.

B. Lossless porous medium equations

$$\partial_i \langle \sigma \rangle - m^{ff} \partial_t \langle v_i^f \rangle - m^{fs} \partial_t \langle v_i^s \rangle = -\langle f_i^f \rangle, \tag{3}$$

$$\partial_k \langle v_k^f \rangle - \kappa^{ff} \partial_t \langle \sigma \rangle - \kappa^{fs} \partial_t \frac{\langle \tau_{kk}^s \rangle}{3} = \langle h^f \rangle, \tag{4}$$

and

$$\partial_j \langle \tau_{ij}^s \rangle - m^{ss} \partial_t \langle v_i^s \rangle - m^{sf} \partial_t \langle v_i^f \rangle = -\langle f_i^s \rangle, \tag{5}$$

$$1/2(\partial_j \langle v_i^s \rangle + \partial_i \langle v_j^s \rangle) - \Lambda^{ss} \delta_{ij} \partial_t \langle \tau_{kk}^s \rangle - 2M^{ss} \partial_t \langle \tau_{ij}^s \rangle - \kappa^{sf} \frac{\delta_{ij}}{3} \partial_t \langle \sigma \rangle = \langle h_{ij}^s \rangle, \tag{6}$$

where $\langle \sigma \rangle$ is the volume-averaged fluid traction [Pa], $\langle v_k^f \rangle$ is the volume-averaged fluid particle velocity [m/s], $\langle \tau_{ij}^s \rangle$ is the volume-averaged solid stress [Pa], $\langle v_k^s \rangle$ is the volume-averaged solid particle velocity [m/s], $\{m^{sf}, m^{fs}\}$ are the coupled mass densities of the fluid/porous composite [kg/m^3], $\{\kappa^{sf}, \kappa^{fs}\}$ are the coupled compressibilities of the fluid/porous composite [Pa^{-1}], and $\{\Lambda^{ss}, M^{ss}\}$ are the coupled compliance coefficients of the fluid/porous composite [Pa^{-1}]. Equations (3)–(6) constitute the basic field equations that pertain to a linear, isotropic, and homogeneous fluid-saturated porous medium, which is lossless in its acoustic behavior[1–3] and which have the same appearance as the high-frequency limit Biot[4–6] equations but a different physical interpretation.[3] Further we have the additional relations

$$m^{ff} = \rho^f - m^{sf}, \tag{7}$$

$$m^{ss} = \rho^s - m^{fs}, \tag{8}$$

$$\kappa^{ff} = \kappa^f - \kappa^{sf}, \tag{9}$$

$$\kappa^{ss} = \kappa^s - \kappa^{fs}, \tag{10}$$

where

$$\kappa^{ss} = 3(3\Lambda^{ss} + 2M^{ss}) = \frac{3}{3\lambda^{ss} + 2\mu^{ss}}, \tag{11}$$

and

$$M^{ss} = \frac{1}{4\mu^{ss}}. \tag{12}$$

Note that ϱ^f is the omnidirectional fluid mass density of the constituting fluid phase [kg/m^3], ϱ^s is the omnidirectional solid mass density of the constituting solid phase [kg/m^3], κ^f is the compressibility of the constituting fluid phase [Pa^{-1}], κ^s is the compressibility of the constituting solid phase [Pa^{-1}], and $\{\lambda^{ss}, \mu^{ss}\}$ are the coupled stiffness coefficients of the fluid/porous composite [Pa]. Moreover, if the porous medium under consideration is reciprocal, we have the additional relations

$$m^{sf} = m^{fs}, \tag{13}$$

$$\kappa^{sf} = \kappa^{fs}. \tag{14}$$

In the following we will assume the porous medium to be reciprocal.

C. Perfectly elastic solid equations

The perfectly elastic solid equations follow directly from the porous medium equations [cf. Eqs. (3)–(6)] by ignoring the fluid parts [Eqs. (3) and (4)] and requiring

$$m^{sf} = m^{fs} = 0, \tag{15}$$

$$\kappa^{sf} = \kappa^{fs} = 0, \tag{16}$$

and

$$\{\Lambda^{ss}, \lambda^{ss}\} = \{\Lambda, \lambda\}, \tag{17}$$

$$\{M^{ss},\mu^{ss}\}=\{M,\mu\}, \tag{18}$$

where λ and μ are the Lamé coefficients of the solid phase [Pa].

III. THE FIELD EQUATIONS IN A MATRIX FORMALISM

To take advantage of the geometry of the configuration the field equations are subjected to two integral transformations: one with respect to time (Laplace transformation), and one with respect to the ''horizontal'' space coordinates x_1, x_2 (Fourier transformation). First we introduce the one-sided Laplace transformation with respect to time. To show the notation we give the transformation equation for the acoustic traction

$$\hat{\sigma}(\mathbf{x},s)=\int_{t=0}^{\infty}\exp(-st)\sigma(\mathbf{x},t)dt, \tag{19}$$

where s is taken to be real and positive. The usefulness of this one-sided Laplace transformation is directly related to the time invariance of the configuration. Further, to take advantage of the shift invariance of the configuration in the ''horizontal'' directions we introduce the Fourier transformation with respect to the spatial horizontal coordinates, x_1 and x_2. To show the notation we give the transformation equation for the acoustic traction:

$$\tilde{\sigma}(\alpha_1,\alpha_2,x_3,s)=\int_{x_1=-\infty}^{\infty}\int_{x_2=-\infty}^{\infty}\exp(is\alpha_\mu x_\mu)\times\hat{\sigma}(\mathbf{x},s)dx_1\,dx_2, \tag{20}$$

where Greek lowercase subscripts are used to indicate the ''horizontal'' parts of tensors, i.e., they are to be assigned the values 1 and 2 only. In Eq. (20), i represents the imaginary unit, and the Fourier-transform parameters α_μ are taken to be real. Note that in terms of the standard Fourier transformation, $s\alpha_\mu$ are the Fourier-transform parameters. As a consequence, the inverse Fourier transformation is given by

$$\hat{\sigma}(\mathbf{x},s)=\left(\frac{s}{2\pi}\right)^2\int_{\alpha_1=-\infty}^{\infty}\int_{\alpha_2=-\infty}^{\infty}\exp(-is\alpha_\mu x_\mu)\times\tilde{\sigma}(\alpha_1,\alpha_2,x_3,s)d\alpha_1\,d\alpha_2. \tag{21}$$

Next, we will subject the relevant field equations to the integral transformations Eqs. (19) and (20). The resulting transformed field equations are put in a matrix formalism. It will show that the mathematical appearance of the matrix formalism is essentially the same for an ideal (frictionless) fluid, a perfectly elastic solid and a lossless porous medium.[1,2] Only the dimensions of the relevant vectors and matrices are different.

A. The ideal (frictionless) fluid

Using the properties $\partial_t\rightarrow s$ and $\partial_\mu\rightarrow -is\alpha_\mu$, according to the integral transformations [Eqs. (19) and (20)], the ideal (frictionless) fluid equations may be put in the following matrix formalism:

$$\partial_3\mathbf{F}+s\mathbf{A}\mathbf{F}=\mathbf{S}, \tag{22}$$

where

$$\mathbf{F}=(-\tilde{\sigma},\tilde{\nu}_3^f)^T, \tag{23}$$

$$\mathbf{A}=\begin{pmatrix}0 & \rho^f\\ \dfrac{\alpha_\mu\alpha_\mu}{\rho^f}+\kappa & 0\end{pmatrix}, \tag{24}$$

$$\mathbf{S}=\left(\tilde{f}_3^f,\tilde{h}^f+\frac{i\alpha_\mu\tilde{f}_\mu^f}{\rho^f}\right)^T. \tag{25}$$

Equation (22) represents the transform-domain matrix formalism for the acoustic wave field vector $\mathbf{F}$ of an ideal (frictionless) fluid of infinite extent, in which sources $\{f_i^f,h^f\}$ are incorporated. The $\mathbf{F}$ vector only contains those field quantities that should be continuous across the horizontal interfaces $x_3=0$, and $x_3=-D$.

B. The lossless porous medium

We first subject Eqs. (3)–(6), in absence of source terms, to a *Laplace* transformation with respect to time and a *Fourier* transformations with respect to x_1 and x_2. Subsequently, we start a process of elimination in such a way that only those quantities remain that should be continuous across the macroscopic interface.[1] In this way we arrive at the following matrix formalism:

$$\partial_3\mathbf{F}^p+s\mathbf{A}^p\mathbf{F}^p=\mathbf{0}, \tag{26}$$

where $\mathbf{A}^p$ is the 8 by 8 acoustic system's matrix for the porous medium and $\mathbf{F}^p$ the porous medium motion-stress vector. The latter one is chosen as the 8 by 1 matrix

$$\mathbf{F}^p=(-\langle\tilde{\sigma}^f\rangle,\ -\langle\tilde{\tau}_{13}^s\rangle,\ -\langle\tilde{\tau}_{23}^s\rangle,\ -\langle\tilde{\tau}_{33}^s\rangle,\ \langle\tilde{\nu}_3^f\rangle,\ \langle\tilde{\nu}_1^s\rangle,\ \langle\tilde{\nu}_2^s\rangle,\ \langle\tilde{\nu}_3^s\rangle)^T, \tag{27}$$

and the expression for the square system's matrix is obtained as[1]

$$\mathbf{A}^{\mathbf{p}}=\begin{pmatrix}\mathbf{R}^{\mathbf{p},T} & \mathbf{S}^{\mathbf{p}}\\ \mathbf{C}^{\mathbf{p}} & \mathbf{R}^{\mathbf{p}}\end{pmatrix}, \tag{28}$$

in which

$$\mathbf{R}^{p,T}=\begin{pmatrix} 0 & 0 & 0 & 0 \\ \left[\frac{m^{sf}}{m^{ff}}+\frac{2\mu^{ss}\kappa^{sf}}{(\lambda^{ss}+2\mu^{ss})\kappa^{ss}}\right] i\alpha_1 & 0 & 0 & \frac{-\lambda^{ss}}{\lambda^{ss}+2\mu^{ss}}\, i\alpha_1 \\ \left[\frac{m^{sf}}{m^{ff}}+\frac{2\mu^{ss}\kappa^{sf}}{(\lambda^{ss}+2\mu^{ss})\kappa^{ss}}\right] i\alpha_2 & 0 & 0 & \frac{-\lambda^{ss}}{\lambda^{ss}+2\mu^{ss}}\, i\alpha_2 \\ 0 & -i\alpha_1 & -i\alpha_2 & 0 \end{pmatrix}, \tag{29}$$

$$\mathbf{S}^{p}=\begin{pmatrix} m^{ff} & 0 & 0 & m^{fs} \\ 0 & S_{11}^{ss} & \frac{3\mu^{ss}}{(\lambda^{ss}+2\mu^{ss})\kappa^{ss}}\,\alpha_1\alpha_2 & 0 \\ 0 & \frac{3\mu^{ss}}{(\lambda^{ss}+2\mu^{ss})\kappa^{ss}}\,\alpha_2\alpha_1 & S_{22}^{ss} & 0 \\ m^{sf} & 0 & 0 & m^{ss} \end{pmatrix}, \tag{30}$$

where

$$S_{11}^{ss}=m^{ss*}+\frac{3\mu^{ss}}{(\lambda^{ss}+2\mu^{ss})\kappa^{ss}}\,\alpha_1\alpha_1+\mu^{ss}\alpha_\sigma\alpha_\sigma, \tag{31}$$

$$S_{22}^{ss}=m^{ss*}+\frac{3\mu^{ss}}{(\lambda^{ss}+2\mu^{ss})\kappa^{ss}}\,\alpha_2\alpha_2+\mu^{ss}\alpha_\sigma\alpha_\sigma, \tag{32}$$

$$m^{ss*}=m^{ss}\left(1-\frac{m^{sf}m^{fs}}{m^{ff}m^{ss}}\right), \tag{33}$$

and

$$\mathbf{C}^{p}=\begin{pmatrix} C^{ff} & 0 & 0 & \frac{\kappa^{fs}}{(\lambda^{ss}+2\mu^{ss})\kappa^{ss}} \\ 0 & \frac{1}{\mu^{ss}} & 0 & 0 \\ 0 & 0 & \frac{1}{\mu^{ss}} & 0 \\ \frac{\kappa^{sf}}{(\lambda^{ss}+2\mu^{ss})\kappa^{ss}} & 0 & 0 & \frac{1}{\lambda^{ss}+2\mu^{ss}} \end{pmatrix}, \tag{34}$$

in which

$$C^{ff}=\kappa^{ff*}+\frac{\kappa^{fs}\kappa^{sf}}{(\lambda^{ss}+2\mu^{ss})\kappa^{ss}\kappa^{ss}}+\frac{1}{m^{ff}}\,\alpha_\sigma\alpha_\sigma, \tag{35}$$

$$\kappa^{ff*}=\kappa^{ff}\left(1-\frac{\kappa^{sf}\kappa^{fs}}{\kappa^{ff}\kappa^{ss}}\right). \tag{36}$$

C. The perfectly elastic solid

The matrix formalism for a perfectly elastic solid follows directly from the lossless porous medium matrix formalism [Eq. (26)]. First one has to omit those field quantities in the wave field vector [Eq. (27)] that pertain to the fluid phase of the porous medium. As a consequence one has to omit the corresponding rows and columns of the $\mathbf{A}^p$ matrix [Eq. (28)]. In this way we arrive at the following matrix formalism:

$$\partial_3\mathbf{F}^s+s\mathbf{A}^s\mathbf{F}^s=\mathbf{0}, \tag{37}$$

where $\mathbf{A}^s$ is the 6 by 6 acoustic system's matrix for the perfectly elastic solid and $\mathbf{F}^s$ is the perfectly elastic solid motion-stress vector. The latter one is chosen as the 6 by 1 matrix

$$\mathbf{F}^s=(-\tilde{\tau}_{13}^s,\ -\tilde{\tau}_{23}^s,\ -\tilde{\tau}_{33}^s,\ \tilde{\nu}_1^s,\ \tilde{\nu}_2^s,\ \tilde{\nu}_3^s)^T, \tag{38}$$

and the expression for the square system's matrix is obtained as

$$\mathbf{A}^s=\begin{pmatrix} \mathbf{R}^{s,T} & \mathbf{S}^s \\ \mathbf{C}^s & \mathbf{R}^s \end{pmatrix}, \tag{39}$$

in which $\mathbf{R}^{s,T}$, $\mathbf{S}^s$, $\mathbf{C}^s$, and $\mathbf{R}^s$ may be found by inspection of Eqs. (28)–(36), together with the identities Eqs. (15)–(18). Next we consider the solution of systems of differential equations of the type of Eqs. (22), (26), and (37). In doing so the configuration will be split in three domains. The first contains the upper fluid half-space ($x_3>0$), the second contains the transmission object ($-D<x_3<0$), and the third contains the lower fluid half-space ($x_3<-D$).

D. The upper fluid half-space ($x_3>0$)

In the space-time domain as well as in the transform domain the total wave field vector in the upper fluid half-space is a superposition of the *incident* wave field vector and the *reflected* wave field vector (at the interface, $x_3=0$). Consequently we may write

$$\mathbf{F}=\mathbf{F}^i+\mathbf{F}^r, \tag{40}$$

where the incident wave field vector, $\mathbf{F}^i$, represents the wave field that is emitted by the sources in an unbounded, homogeneous, and isotropic ideal (frictionless) fluid, while the reflected wave field vector, $\mathbf{F}^r$, is represented by part of the

homogeneous solution of Eq. (22), i.e., that part that vanishes if $x_3 \to \infty$.

1. Incident wave field

Taking into account the fact that the source located at $(0,0,h^T)$ is a monopole point source (volume-injection type), the incident wave field vector follows from [cf. Eq. (22)]:

$$\partial_3 \mathbf{F}^i + s\mathbf{A}\mathbf{F}^i = \mathbf{S}'\,\delta(x_3 - h^T), \tag{41}$$

where

$$\mathbf{S}' = (0,\hat{Q})^T, \tag{42}$$

in which

$$\hat{Q} = \hat{Q}(s) = \mathscr{L}\{Q(t)\}. \tag{43}$$

In Eq. (42), $Q(t)$ is representative of the source signature. Because Eq. (41) is a partial differential equation in terms of x_3 the wave field vector $\mathbf{F}^i$ should be a continuous function of x_3. Consequently we may integrate Eq. (41) with respect to x_3 over the interval $h^T - \delta \leqslant x_3 \leqslant h^T + \delta$. Upon taking the limit, $\delta \to 0$, and requiring the continuity of the wave field vector, we arrive at the condition

$$\lim_{x_3 \downarrow h^T} \mathbf{F}^i - \lim_{x_3 \uparrow h^T} \mathbf{F}^i = \mathbf{S}'. \tag{44}$$

Next we focus on the solution of the homogeneous version of Eq. (41). From elementary matrix theory we know that if a $N \times N$ matrix $\mathbf{A}$ has N independent eigen vectors there exists a $N \times N$ transformation matrix $\mathbf{T}$ in such a way that $\mathbf{T}$ comprises the eigen vectors of $\mathbf{A}$ as column vectors. Then $\mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ is a diagonal matrix with the eigenvalues of $\mathbf{A}$ on its diagonal. Next we apply the transformation

$$\mathbf{F}^i = \mathbf{T}\mathbf{F}^{i'} \tag{45}$$

to the homogeneous version of Eq. (41) and multiply the result with $\mathbf{T}^{-1}$. This results in

$$\partial_3 \mathbf{F}^{i'} + s[\mathbf{T}^{-1}\mathbf{A}\mathbf{T}]\mathbf{F}^{i'} = \mathbf{0}, \tag{46}$$

where

$$\mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \begin{pmatrix} \gamma & 0 \\ 0 & -\gamma \end{pmatrix}. \tag{47}$$

Equation (46) is the diagonalized form of the homogeneous version of Eq. (41) and has the general solution

$$F_1^{i'} = C_1^{\mathrm{UP}} \exp(-s\gamma x_3), \tag{48}$$

$$F_2^{i'} = C_2^{\mathrm{DOWN}} \exp(s\gamma x_3), \tag{49}$$

where

$$\gamma = \left(\frac{1}{c_f^2} + \alpha_\mu \alpha_\mu\right)^{1/2} \qquad \mathrm{Re}(\gamma) > 0. \tag{50}$$

In Eq. (50) c_f represents the water wave speed, i.e., $c_f = (\varrho^f \kappa)^{-1/2}$. According to Eq. (45) we have

$$F_1^i = \frac{\rho^f}{\gamma}\,[C_1^{\mathrm{UP}} \exp(-s\gamma x_3) - C_2^{\mathrm{DOWN}} \exp(s\gamma x_3)], \tag{51}$$

$$F_2^i = C_1^{\mathrm{UP}} \exp(-s\gamma x_3) + C_2^{\mathrm{DOWN}} \exp(s\gamma x_3). \tag{52}$$

When we apply the jump condition, Eq. (44), to Eqs. (51) and (52) the coefficients, C_1^{UP} and C_2^{DOWN}, can be determined. From this result, the incident wave field follows as

$$\begin{aligned} F_1^i &= \frac{\rho^f}{2\gamma}\,\hat{Q}(s)\exp(-s\gamma|x_3 - h^T|), \\ F_2^i &= \tfrac{1}{2}\,\mathrm{sgn}(x_3 - h^T)\hat{Q}(s)\exp(-s\gamma|x_3 - h^T|) \quad x_3 \gtrless h^T. \end{aligned} \tag{53}$$

2. The reflected wave field

The reflected wave field vector, $\mathbf{F}^r$, follows from the homogeneous version of Eq. (22), confining the solution to that part that remains bounded when $x_3 \to \infty$. Therefore the solution for the reflected wave field vector follows from inspection of Eqs. (51) and (52) as

$$\begin{pmatrix} F_1^r \\ F_2^r \end{pmatrix} = 1/2 \begin{pmatrix} \dfrac{\rho^f}{\gamma} \\ 1 \end{pmatrix} R^F \exp(-s\gamma[x_3 + h^T]), \tag{54}$$

where the factor $\exp(-s\gamma h^T)/2$ has been added for convenience. R^F represents the reflection coefficient at the upper fluid/solid (fluid/porous) interface ($x_3 = 0$), respectively.

3. The total wave field for the upper fluid half-space

The total wave field vector for the upper fluid half-space follows, according to the superposition principle, from Eqs. (40), (53), and (54) as

$$\begin{aligned} -\tilde{\sigma}(x_3) = \frac{\rho^f}{2\gamma}\,[&\hat{Q}(s)\exp(-s\gamma|x_3 - h^T|) \\ &+ R^F \exp(-s\gamma(x_3 + h^T)], \end{aligned} \tag{55}$$

$$\begin{aligned} \tilde{v}_3^f(x_3) = 1/2[&\mathrm{sgn}(x_3 - h^T)\hat{Q}(s)\exp(-s\gamma|x_3 - h^T|) \\ &+ R^F \exp(-s\gamma(x_3 + h^T))]; \quad x_3 \gtrless h^T. \end{aligned} \tag{56}$$

E. The transmission object ($-D < x_3 < 0$)

In this section we will give the solution for the wave field vector $\mathbf{F}^p$ and $\mathbf{F}^s$ for, respectively, the lossless porous medium [cf. Eq. (26)] and the perfectly elastic solid [cf. Eq. (37)]. In solving these equations we proceed in exactly the same way as we did in solving the homogeneous version of Eq. (22). The eigenvector matrices $\mathbf{T}^p$ and $\mathbf{T}^s$ are, respectively, given in Refs. 1 and 2.

1. The lossless porous medium

Applying the transformation

$$\mathbf{F}^p = \mathbf{T}^p \mathbf{F}^{p'}, \tag{57}$$

Eq. (26) is transformed into

$$\partial_3 \mathbf{F}^{p'} + s[(\mathbf{T}^p)^{-1}\mathbf{A}^p\mathbf{T}^p]\mathbf{F}^{p'} = \mathbf{0}, \tag{58}$$

where

$$(\mathbf{T}^p)^{-1}\mathbf{A}^p\mathbf{T}^p = \begin{pmatrix} -\mathbf{D}^p & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^p \end{pmatrix}, \tag{59}$$

and in which

$$\mathbf{D}^p = \begin{pmatrix} \gamma^+ & 0 & 0 & 0 \\ 0 & \gamma^- & 0 & 0 \\ 0 & 0 & \gamma^{\mathrm{sh}} & 0 \\ 0 & 0 & 0 & \gamma^{\mathrm{sv}} \end{pmatrix}, \tag{60}$$

where

$$\gamma^{+,-,\mathrm{sh},\mathrm{sv}} = \left(\frac{1}{c^2_{+,-,\mathrm{sh},\mathrm{sv}}} + \alpha_\mu \alpha_\mu \right)^{1/2} \quad \mathrm{Re}(\gamma^{+,-,\mathrm{sh},\mathrm{sv}}) > 0. \tag{61}$$

The quantities γ^+, γ^-, γ^{sh}, and γ^{sv} denote the vertical slownesses of the fast compressional wave, slow compressional wave, shear wave with horizontal polarization (with respect to the interface), and the shear wave with vertical polarization (with respect to the interface), respectively. Solving Eq. (58) and applying the transformation Eq. (57) we arrive at the following expressions for the wave field vector $\mathbf{F}^p$

$$F_i^p = 1/2 T_{ik}^p C_k^{\mathrm{UP}} \exp[-s(\gamma^{(k)} x_3 + \gamma h^T)] + 1/2 T_{il}^p C_l^{\mathrm{DOWN}} \exp[s(\gamma^{(l)} x_3 - \gamma h^T)], \tag{62}$$

where

$$i = 1 \cdots 8, \quad k = 5 \cdots 8, \quad l = 1 \cdots 4,$$
$$\{\gamma^{(1)}, \gamma^{(2)}, \gamma^{(3)}, \gamma^{(4)}\} = \{\gamma^{(5)}, \gamma^{(6)}, \gamma^{(7)}, \gamma^{(8)}\} = \{\gamma^+, \gamma^-, \gamma^{\mathrm{sv}}, \gamma^{\mathrm{sh}}\}. \tag{63}$$

In Eq. (62) the factor, $\frac{1}{2}\exp(-s\gamma h^T)$, is added for convenience and the elements T_{ij}^p are the constituents of $\mathbf{T}^p$.[1,2]

2. The perfectly elastic medium

Similarly one may arrive at the expression for the perfectly elastic wave field vector $\mathbf{F}^s$. It can be written as [cf. Eq. (13)]

$$F_i^s = 1/2 T_{ik}^s C_k^{\mathrm{DOWN}} \exp(s[\gamma^{(k)} x_3 - \gamma h^T]) + 1/2 T_{il}^s C_l^{\mathrm{UP}} \exp(-s[\gamma^{(l)} x_3 + \gamma h^T]), \tag{64}$$

where

$$i = 1 \cdots 6, \quad k = 1 \cdots 3, \quad l = 4 \cdots 6,$$
$$\{\gamma^{(1)}, \gamma^{(2)}, \gamma^{(3)}\} = \{\gamma^{(4)}, \gamma^{(5)}, \gamma^{(6)}\} = \{\gamma^p, \gamma^{\mathrm{sv}}, \gamma^{\mathrm{sh}}\}. \tag{65}$$

Again, in Eq. (64) the factor, $\frac{1}{2}\exp(-s\gamma h^T)$, is added for convenience and the elements T_{ij}^p are the constituents of $\mathbf{T}^s$.[2]

It can be observed from Eqs. (62) and (64) that the total wave field in the transformed domain consists of UP and DOWN going waves. The coefficient vectors $\mathbf{C}^{\mathrm{UP}}$ and $\mathbf{C}^{\mathrm{DOWN}}$, respectively, represent the reflection coefficients at the lower and upper fluid/porous or fluid/elastic solid interface. They follow from the boundary conditions.

F. The lower fluid half-space ($x_3 < -D$)

In the lower fluid half-space no sources are present and consequently the solution for the wave field vector $\mathbf{F}$ in the lower fluid half-space follows from the homogeneous solution of Eq. (22), confining the solution to that part that vanishes for $x_3 \to -\infty$. Consequently, the wave field vector $\mathbf{F}$ for the lower fluid half-space may be written down by inspection of Eqs. (51) and (52)

$$-\tilde{\sigma}(x_3) = -\frac{\rho^f}{2\gamma} T^F \exp(s\gamma[x_3 - h^T]), \tag{66}$$

$$\tilde{v}_3^f(x_3) = 1/2 T^F \exp(s\gamma[x_3 - h^T]), \tag{67}$$

where T^F represents the transmission coefficient at the lower fluid/elastic solid or fluid/porous interface ($x_3 = -D$), respectively. The factor $\frac{1}{2}\exp(-s\gamma h^T)$ has been added for convenience.

IV. BOUNDARY CONDITIONS AT THE FLUID/POROUS INTERFACES

In our new poroelastic theory, a procedure of operator decomposition applied on both the relevant field equations for the ideal (frictionless) fluid and the fluid-saturated porous medium results in the following transformed domain boundary conditions at both the fluid/porous interfaces:[1,2]

$$\langle \tilde{\sigma} \rangle = \tilde{\sigma}, \quad \langle \tilde{v}_3^f \rangle = \tilde{v}_3^f, \quad \langle \tilde{\tau}_{3k}^s \rangle = 0. \tag{68}$$

In contrast to these boundary conditions we mention here the boundary conditions often used for the Biot theory, according to Deresiewicz and Skalak:[14]

$$\langle \tilde{\sigma} \rangle = \phi \tilde{\sigma}, \quad \langle \tilde{v}_3^f \rangle + \langle \tilde{v}_3^s \rangle = \tilde{v}_3^f, \quad \langle \tilde{\sigma} \rangle + \langle \tilde{\tau}_{3k}^s \rangle = \tilde{\sigma}. \tag{69}$$

Note that Eq. (69) is formulated for free flow through the surface pores at the boundary between the fluid and the porous medium. Furthermore, all quantities are total volume averaged, according to the conventions used by De Vries[1] and Geerits.[3] In this respect, the Biot theory shows an inconsistency. Biot defined the macroscopic particle velocities and the acoustic pressure to be intrinsic volume averaged ones and the macroscopic stresses to be total volume averaged ones. This explains the porosity quantity in the boundary condition for the fluid and stress.

A. Upper fluid/porous interface: ($x_3 = 0$)

Applying the boundary conditions Eq. (68) to the upper fluid/porous interface, it follows from Eqs. (55), (56), and (62) that

$$\left(T_{1k}^p - \frac{\rho^f}{\gamma} T_{5k}^p \right) C_k^{\mathrm{UP}} + \left(T_{1l}^p - \frac{\rho^f}{\gamma} T_{5l}^p \right) C_l^{\mathrm{DOWN}} = \frac{2\rho^f}{\gamma} \hat{Q}(s), \tag{70}$$

$$T_{ik}^p C_k^{\mathrm{UP}} + T_{il}^p C_l^{\mathrm{DOWN}} = 0, \quad i = 2,3,4. \tag{71}$$

In writing down Eq. (70) we have eliminated the reflection coefficient R^F from the first and second boundary condition [cf. Eq. (68)].

B. Lower fluid/porous interface: ($x_3 = -D$)

Applying the boundary conditions Eq. (68) to the lower fluid/porous interface, it follows from Eqs. (66), (67), and (62) that

$$\left(T^p_{1k}+\frac{\rho^f}{\gamma}T^p_{5k}\right)C^{\mathrm{UP}}_k\exp(s\gamma^{(k)}D)+\left(T^p_{1l}+\frac{\rho^f}{\gamma}T^p_{5l}\right)C^{\mathrm{DOWN}}_l \times\exp(-s\gamma^{(l)}D)=0, \tag{72}$$

$$T^p_{ik}C^{\mathrm{UP}}_k\exp(s\gamma^{(k)}D)+T^p_{il}C^{\mathrm{DOWN}}_l\exp(-s\gamma^{(l)}D)=0, \quad i=2,3,4. \tag{73}$$

In writing down Eq. (72) we have eliminated the transmission coefficient, T^F, from the first and second boundary condition [cf. Eq. (68)]. The transmission coefficient follows as

$$T^F=T^p_{5k}C^{\mathrm{UP}}_k\exp[s(\gamma^{(k)}+\gamma)D] + T^p_{5l}C^{\mathrm{DOWN}}_l\exp[-s(\gamma^{(l)}-\gamma)D]. \tag{74}$$

We now have a system of eight equations [Eqs. (70)–(73)] with eight unknown coefficients. However, because horizontally polarized shear waves cannot be transmitted to the lower fluid half-space, it follows from Eqs. (70)–(73) that

$$C^{\mathrm{DOWN}}_4=C^{\mathrm{UP}}_8. \tag{75}$$

Substituting Eq. (75) in Eqs. (70)–(73) results in the following matrix formalism:

$$\mathbf{B_1}\mathbf{C}^{\mathrm{UP}}+\mathbf{B_2}\mathbf{C}^{\mathrm{DOWN}}=\mathbf{b}', \tag{76}$$

$$\mathbf{B_3}\mathbf{E^{-1}}\mathbf{C}^{\mathrm{UP}}+\mathbf{B_4}\mathbf{E}\mathbf{C}^{\mathrm{DOWN}}=\mathbf{0}, \tag{77}$$

where

$$\mathbf{C}^{\mathrm{DOWN}}=(C^{\mathrm{DOWN}}_1,\ C^{\mathrm{DOWN}}_2,\ C^{\mathrm{DOWN}}_3)^T, \tag{78}$$

and

$$\mathbf{C}^{\mathrm{UP}}=(C^{\mathrm{UP}}_5,\ C^{\mathrm{UP}}_6,\ C^{\mathrm{UP}}_7)^T. \tag{79}$$

Further $\mathbf{E}$ is given by

$$\mathbf{E}=\begin{pmatrix}\exp(-s\gamma^+D) & 0 & 0\\ 0 & \exp(-s\gamma^-D) & 0\\ 0 & 0 & \exp(-s\gamma^{\mathrm{sv}}D)\end{pmatrix}. \tag{80}$$

The matrices $\mathbf{B}_1$, $\mathbf{B}_2$, $\mathbf{B}_3$, and $\mathbf{B}_4$ form a subset from the eigenvector matrix and are given in Ref. 2. Vector $\mathbf{b}'$ is also presented in Ref. 2. From Eqs. (76) and (77) it follows that

$$\mathbf{C}^{\mathrm{UP}}=-(\mathbf{E}\mathbf{B_3^{-1}}\mathbf{B_4}\mathbf{E})\mathbf{C}^{\mathrm{DOWN}}=-\mathbf{M}\mathbf{C}^{\mathrm{DOWN}}, \tag{81}$$

and

$$\mathbf{C}^{\mathrm{DOWN}}=[\mathbf{I}-(\mathbf{B_2^{-1}}\mathbf{B_1})\mathbf{M}]^{-1}\mathbf{B_2^{-1}}\mathbf{b}'. \tag{82}$$

To distinguish between the primaries and the multiples we apply a Neumann expansion to Eq. (82). This results in

$$\mathbf{C}^{\mathrm{DOWN}}=[\mathbf{I}+\mathbf{X}^1+\mathbf{X}^2+\mathbf{X}^3+\cdots+\mathbf{X}^n]\mathbf{B_2^{-1}}\mathbf{b}', \tag{83}$$

where n denotes the number of multiples and in which

$$\mathbf{X}=(\mathbf{B_2^{-1}}\mathbf{B_1})\mathbf{M}=(\mathbf{B_2^{-1}}\mathbf{B_1})\mathbf{E}(\mathbf{B_3^{-1}}\mathbf{B_4})\mathbf{E}. \tag{84}$$

Note that the Neumann expansion [Eq. (83)] is an essential step in order to apply the *Cagniard–de Hoop* method. When taking no multiples into account, Eq. (83) simplifies to

$$\mathbf{C}^{\mathrm{DOWN}}=\mathbf{B_2^{-1}}\mathbf{b}'. \tag{85}$$

Substituting Eqs. (81) and (85) in Eq. (74) yields

$$T^F=\hat{Q}(s)\exp(sD\gamma)\sum_j A^j\exp(-sD\gamma^j), \tag{86}$$

where j only takes on the values 1, 2, 3, and A^1, A^2, A^3 represent the amplitude factors of the fast P, slow P, and S wave, respectively.[2]

V. BOUNDARY CONDITIONS AT THE FLUID/ELASTIC SOLID INTERFACES

In the transformed domain the boundary conditions at both fluid/elastic solid interfaces follow as

$$\tilde{\tau}^s_{13}=0,\quad \tilde{\tau}^s_{23}=0,\quad \tilde{\tau}^s_{33}=\tilde{\sigma}^f,\quad \tilde{\nu}^s_3=\tilde{\nu}^f_3. \tag{87}$$

The matrix formalism occurring due to the boundary conditions is exactly the same as for the porous medium [Eqs. (76) and (77)]. The only differences are the dimensions and the contents of the relevant matrices. For the perfectly elastic medium, the relevant matrices (T^s) and (B_1,B_2,B_3,B_4) are presented in.[2] Furthermore, Eq. (86) remains valid for a perfectly elastic medium, where j takes on the values 1 and 2 only, and where A^1, A^2 represent the amplitude factors of the P- and S-wave, respectively. They are presented in Ref. 2.

VI. THE *CAGNIARD–DE HOOP* METHOD

In this section we will give a expression for the acoustic traction in the lower fluid half-space ($x_3<-D$), in the $(\mathbf{x},t)$ domain. To this end we substitute Eq. (86) in Eq. (66) and apply the back transformation, according to Eq. (21). This yields

$$\hat{\sigma}(\mathbf{x}^R,s)=s^2\hat{Q}(s)\hat{G}(\mathbf{x}^R,s), \tag{88}$$

where

$$\hat{G}(\mathbf{x}^R,s)=(2\pi)^{-2}\int_{\alpha_1=-\infty}^{\infty}\int_{\alpha_2=-\infty}^{\infty}\Pi^{(j)}(\alpha_1\alpha_2) \times\exp(-s[i\alpha_\mu x^R_\mu+h_\lambda\gamma_\lambda])d\alpha_1\,d\alpha_2 \tag{89}$$

is by definition the Green's function (response function). Further we have

$$\Pi^{(j)}(\alpha_1,\alpha_2)=\frac{\rho^f}{2\gamma}A^j, \tag{90}$$

and

$$h_1=h^T-(x^R_3+D),\quad h_2=D, \qquad \gamma_1=\gamma,\quad \gamma_2=\gamma_j, \tag{91}$$

where j relates to the P- or S-wave component of the Green's function. The aim of the *Cagniard–de Hoop* technique is now to rewrite Eq. (89), in such a way, that it takes the form

$$\hat{G}(\mathbf{x}^R,s)=\int_{\tau=T}^{\infty}\exp(-s\tau)W^G(\mathbf{x}^R,\tau)d\tau, \tag{92}$$

where τ is a real variable of integration, $T>0$, $\mathrm{Re}\{s\}>0$ and where $W^G(\mathbf{x},\tau)$ does not depend on s. Suppose that this has been achieved, then the uniqueness theorem of the Laplace

transformation with real, positive, transform parameter (Lerch's theorem[12]) ensures that

$$G(\mathbf{x}^R,t)=\begin{cases}0 & t<T\\ W^G(\mathbf{x}^R,t) & t>T.\end{cases} \tag{93}$$

Additional properties of the Laplace transformation then lead to the final result in the space-time domain

$$\sigma(\mathbf{x}^R,t)=\begin{cases}0 & t<T\\ \partial_t^2\int_{\tau=T}^{t} Q(t-\tau)W^G(\mathbf{x}^R,t)\,d\tau & t>T,\end{cases} \tag{94}$$

in which Q is a representative of the source signature [cf. Eq. (43)]. From Eq. (94) it is evident that T is the arrival time of the generalized ray constituent (P- or S-wave) under consideration. In rewriting Eq. (89) in the form of Eq. (92) several steps are carried out; they are characteristic for the *Cagniard–de Hoop* technique.[17,18] The version of the Cagniard–de Hoop technique we employ uses the transformations

$$\alpha_1=-ip\,\cos\,\phi-q\,\sin\,\phi, \tag{95}$$

$$\alpha_2=-ip\,\sin\,\phi+q\,\cos\,\phi, \tag{96}$$

where

$$p\in I,\quad q\in \mathsf{R},\quad 0\leqslant\phi<2\pi, \tag{97}$$

and where ϕ follows from the polar-coordinate specification of the point of observation in the x_1-x_2 plane

$$x_1^R=r\,\cos\,\phi, \tag{98}$$

$$x_2^R=r\,\sin\,\phi. \tag{99}$$

Substituting Eqs. (95)–(99) in Eq. (89) results in

$$\hat{G}(\mathbf{x}^R,s)=(4\pi^2 i)^{-1}\int_{q=-\infty}^{\infty}dq\int_{p=-i\infty}^{i\infty}\overline{\Pi}(p,q)\times\exp(-s[pr+h_\lambda\overline{\gamma_\lambda}(p,q)])dp, \tag{100}$$

where Π results from $\Pi^{(j)}$ and $\overline{\gamma_\lambda}$ from γ_λ under the substitutions of Eqs. (95)–(99). After deforming the path of integration with respect to p into the complex p plane via the *Cagniard–de Hoop contour*

$$\tau=pr+\overline{\gamma_\lambda}h_\lambda, \tag{101}$$

keeping τ real and positive, we end up with

$$\hat{G}(\mathbf{x}^R,s)=(2\pi^2)^{-1}\int_{\tau=\tau^B(0)}^{\infty}\exp(-s\tau)d\tau\times\int_{q=-Q^B(\tau)}^{Q^B(\tau)}\mathrm{Im}\left[\overline{\Pi}(\tau,q)\,\frac{\partial p^B(\tau,q)}{\partial\tau}\right]dq, \tag{102}$$

in which $q=Q^B(\tau)$ is the inverse mapping of $\tau=\tau^B(q)$. Note that $\tau^B(q)$ can be found by substituting $p=p^B(q)$ (intersection of *Cagniard–de Hoop* contour with real $\underline{p}$ axis, which follows from $\partial\tau/\partial p=0$) in Eq. (101) and that $\overline{\Pi}(\tau,q)$ results from the substitution $p=p^B(\tau,q)$ in $\overline{\Pi}(p,q)$. For more detail the reader is referred to Ref. 2 (pp. 106–120) and Refs. 17 and 19. From Eq. (102) it can be deduced that $\tau=\tau^B(0)$ represents the arrival time of the body wave (P wave or S wave). Comparing Eq. (102) with Eq. (92), it follows from Eq. (94) that the acoustic traction in the space-time domain, if only a body wave is present, may be written as

$$\sigma(\mathbf{x}^R,\tau)=(2\pi^2)^{-1}\partial_\tau^2\int_{u=\tau^B(0)}^{\tau}Q(\tau-u)\times\left(\int_{q=-Q^B(u)}^{Q^b(u)}\mathrm{Im}\left[\overline{\Pi}(u,q)\,\frac{\partial p^B(u,q)}{\partial u}\right]dq\right)du;\quad \tau>\tau^B(0). \tag{103}$$

In this paper we assume that no surface waves and/or head waves will occur. Moreover, it is conjectured that if they would occur, their contribution relative to the body wave contribution would be small. The fact that this contribution indeed is small relative to the body wave contribution is proven for the configuration under consideration in Ref. 2.

VII. THEORY VERSUS EXPERIMENTS

Whereas in the numerical experiment the source is taken to be a monopole point source, in the laboratory experiments we use a flat, circular piezoelectric source. Further, in the numerical experiment the receiver is taken to be a notional point receiver, i.e., it does not influence the actual response at the point of observation. In the laboratory experiments, however, the receiver is identical to the source. Consequently we have to investigate under which conditions the piezo-electric transducers, transmitter, and receiver may be replaced by a monopole point source and a notional point receiver, respectively. For this purpose we consider the following two experiments:

(1) Consider a monopole point source (volume-injection type) in a homogeneous and isotropic ideal (frictionless) fluid of infinite extent, with source signature $Q(t)$. The acoustic traction response at an arbitrary point of observation in the $(\mathbf{x},s)$ domain, when only a volume-injection source is taken into account (i.e., f_i^f and $h^f\neq 0$) is very well-known and may be written as

$$\hat{\sigma}(\mathbf{x},s)=-\frac{\rho^f}{4\pi|\mathbf{x}|}\,s\hat{Q}(s)\exp\left(-s\,\frac{|\mathbf{x}|}{c}\right). \tag{104}$$

(2) Consider a pulsating, flat, circular, piezoelectric transducer in a homogeneous and isotropic ideal (frictionless) fluid of infinite extent.

When an alternating voltage is applied to the piezoelectric element, it alternately expands and contracts. In this way electrical energy is converted into mechanical energy. This mechanical energy is transmitted to the fluid. Inversely, a pressure field acting on the piezoelectric element induces an alternating voltage over the piezoelectric element. If we are in the linear regime, the induced voltage over the piezoelectric element is proportional to the applied pressure. The piezo-electric source can be modeled as a moving piston in an infinite, rigid baffle. This situation is identical to a plane wave incident upon a circular aperture of which the wave-

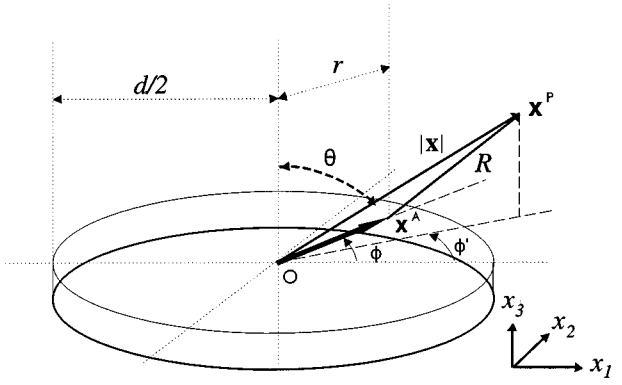


FIG. 2. Contribution of one point source (A) to the point of observation (P).

length (λ) is much smaller then the diameter of the aperture (d). The diameter of the circular aperture equals the diameter of the circular, piezoelectric element.

According to Huygens principle the contributions of all points on the piezo-electric surface to the acoustic traction at the point of observation $\mathbf{x}^p$ (Fig. 2), can be found by the integration of $-(\rho^f/4\pi R)s\hat{Q}(s)\exp(-sR/c_f)$ (the contribution of one point source) over the entire surface. This results in the following expression:

$$\hat{\sigma}(\mathbf{x},s,\Theta)=-\frac{\rho^f}{4\pi|\mathbf{x}|}\,s\hat{Q}(s)\exp\left(-s\,\frac{|\mathbf{x}|}{c}\right)D(\Theta,s), \tag{105}$$

where

$$D(\Theta,s)=\frac{\pi d^2}{2}\,\frac{J_1(-isd\,\sin\,\Theta/2c_f)}{(-isd\,\sin\,\Theta/2c_f)}. \tag{106}$$

In deriving Eqs. (105) and (106) we have assumed $0<r/|\mathbf{x}|\ll 1$, which is representative of the so-called *far-field* approximation. $D(\Theta,s)$ represents the directional characteristic (directivity term) of a flat, circular piezoelectric transducer in the far field. Θ (see Fig. 3) will be called the *offset angle*. The minimum distances for which Eqs. (105) and (106) are valid follow from the empirical relation[21]

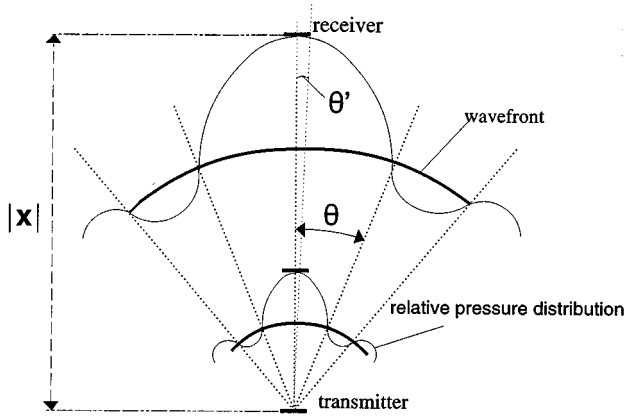


FIG. 3. Relative pressure distribution at increasing distance (f=constant).

$$|\mathbf{x}|>0.75\,\frac{d^2}{\lambda_w}. \tag{107}$$

Next, we focus on the relative pressure distribution at an arbitrary position in the far field. From Eq. (105), it follows that

$$\frac{\hat{\sigma}(\mathbf{x},s,\Theta)}{\lim\limits_{\Theta\to 0}\hat{\sigma}(\mathbf{x},s,\Theta)}=2\,\frac{J_1(-isd\,\sin\,\Theta/2c_f)}{(-isd\,\sin\,\Theta/2c_f)}\,. \tag{108}$$

If we investigate the functional dependence of the absolute value of the relative pressure distribution versus the offset angle Θ, at different frequencies, two important features can be distinguished: First, that on-axis ($\Theta=0$ degrees), in the far field, the relative pressure distribution always equals one. In that case it follows that the acoustic traction response due to one point source [cf. Eq. (104)] is proportional to the acoustic traction field emitted by a flat, circular, piezoelectric transducer [cf. Eq. (105)]. Second, we observe that the fluctuation in the relative pressure field near $\Theta=0$ becomes less pronounced if the frequency is decreased. Due to the spherical character (Θ dependence) of Eq. (106), the same effect is observed if the distance $|\mathbf{x}|$ is increased, while the frequency is kept constant. This becomes evident in Fig. 3. If we perform an experiment as depicted in Fig. 3 in a water tank and we would multiply the amplitude spectra of the measured signals by $|\mathbf{x}|$ one would expect the result to be proportional to $s\hat{Q}(s)$. This follows from Eq. (105) if $\Theta\to 0$ and if we assume the receiver to be of the notional type, i.e., the receiver does not influence the actual response in the point of observation. However, if the receiver is identical to the source (i.e., a flat, circular piezoelectric transducer), the proportionality with $s\hat{Q}(s)$ will only be observed if the pressure distribution on the receiver is *spatially* constant. This, is only true for the limiting case: $|\mathbf{x}|\to\infty$ (see Fig. 3) and approximately if

$$|\mathbf{x}|\geqslant 1.5\,\frac{d^2}{\lambda_w}, \tag{109}$$

i.e., twice the far-field distance [cf. Eq. (107)]. The validity of Eq. (109) followed from experiments. Substituting the characteristic parameters relating to the used transducers ($d=2.86\times10^{-2}$ m, $f=0.5$ MHz, $c_f=1476.9$ m/s) in Eq. (109) yields $|\mathbf{x}|>0.42$ m. At this distance, the angle θ' (see Fig. 3) can be calculated as: $\theta'=0.034$ rad. This is the angle under which the transmitter observes the rim of the receiver. It can be calculated that, even in this case ($\theta'=0.034$ rad), the relative pressure is very close to unity (to be exact; 87% of its maximum). Therefore, if Eq. (109) is satisfied and $\Theta\to 0$, the recalculated measured water signal is proportional to $s\hat{Q}(s)$. If this measured water signal is denoted with $H(t)$, (see Fig. 4), the proportionality may be expressed as

$$\partial_t Q(t)\propto H(t). \tag{110}$$

A typical picture of $H(t)$ and its corresponding amplitude spectrum is presented in Figs. 4 and 5. To show the ''quasi'' plane wave behavior of the emitted pressure field in the far field, we consider two on-axis frequency spectra at different distances from the source, say R_1, R_2, where $R_2>R_1$. Di-

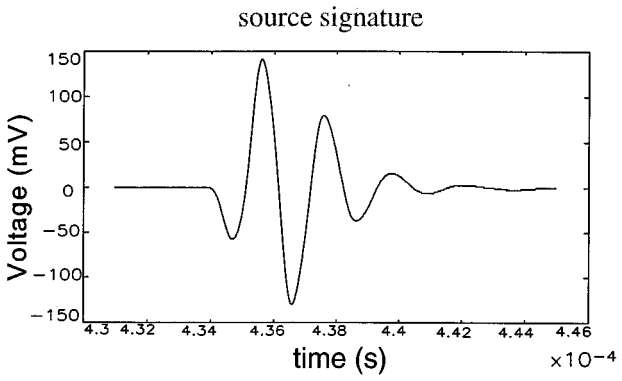


FIG. 4. Source signature ($H(t)$).

viding these two spectra on each other we arrive at

$$\frac{\hat{\sigma}(R_1,s)}{\hat{\sigma}(R_2,s)} \propto \exp\left(\frac{s[R_2-R_1]}{c_f}\right) = \exp\left(\frac{i2\pi f[R_2-R_1]}{c_f}\right), \tag{111}$$

where we used the relation $s=i\omega=i2\pi f$. The phase angle $\phi_p(f)$ [rad] follows from inspection of Eq. (111) as

$$\phi_p(f) = 2\pi\left(\frac{R_2-R_1}{c_f}\right) f. \tag{112}$$

From Eq. (112) it follows that there should be a linear relationship between the phase angle and frequency, i.e., if there is an uniform pressure distribution on the piezoelectric receiver element. In Fig. 6 such a phase spectrum is presented, where $R_2=0.65$ m and $R_1=0.45$ m. Note that $-\pi < \phi_p(f) \leqslant \pi$.

Therefore, when we combine Eq. (103) with Eq. (110) we arrive at

$$p(\mathbf{x}^R,\tau) \propto -\partial_\tau \int_{u=\tau^B(0)}^{\tau} H(\tau-u) \times \left(\int_{q=-Q^B(u)}^{Q^B(u)} \mathrm{Im}\left[\overline{\Pi}(u,q)\, \frac{\partial p^B(u,q)}{\partial u}\right] dq \right) du; \quad \tau > \tau^B(0), \tag{113}$$

where we have replaced the fluid traction by its opposite, i.e., the fluid pressure, which is an intrinsic physical quantity.

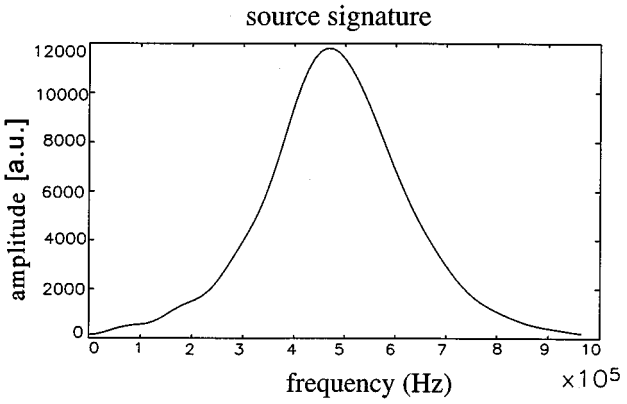


FIG. 5. Amplitude spectrum of source signature.

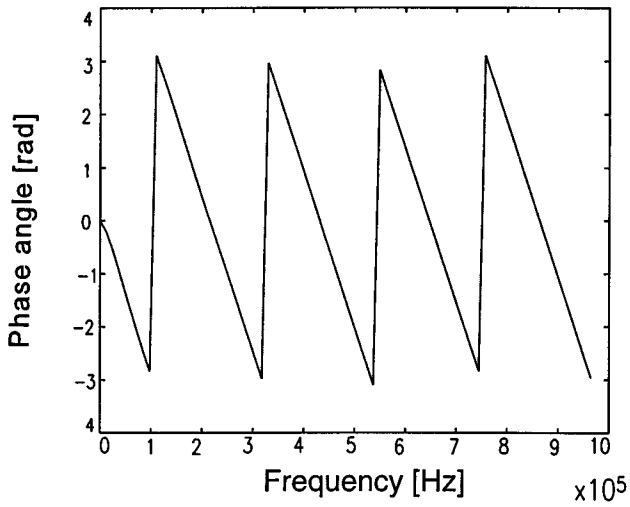


FIG. 6. Phase spectrum of division between two on-axis frequency spectra in a water tank.

Equation (113) will be used to calculate the synthetic seismograms. The proportionality constant suggested by Eq. (113) should be the same for all bulk waves.

In presenting the measured and calculated data we will first present a brief summary of the perfectly elastic solid seismograms and subsequently we will present the porous medium seismograms. The former only serves as the validation of the comparison technique.

A. The perfectly elastic solid (steel)

From the literature it is well-known that there may be considerable discrepancy between the static measured constitutive parameters and their dynamic equivalents, i.e., those occurring in the elasto-dynamic wave field equations. In particular, this is true when the medium under consideration is a porous medium. The static and dynamic densities, however, are almost always the same. Consequently, it was decided to determine the constitutive parameters indirectly from the acoustic measurements by means of the wave speeds. The constitutive parameters compressibility (κ') and shear modulus (μ) follow from

$$\kappa^s = \frac{3}{\rho^s(3c_p^2-4c_s^2)}, \tag{114}$$

$$\mu = \rho^s c_s^2. \tag{115}$$

The input parameters for the numerical model concerning the perfectly elastic medium are presented in Table I. The measurement configuration for the P wave is depicted in Fig. 1. The P-wave measurements were performed on a wave front, i.e., we simulated the incidence of part of a spherical wave front by rotating the transmitter to a set of subsequent angles θ_i (Fig. 1) and we scanned the diffracted wave front in the lower fluid half-space at a particular arrival time. By means of the acoustic wave speeds c_f and c_p, one may calculate the position of a diffracted wave front in the lower fluid half-space at any desired time instant ($T_{\mathrm{ar}}=396.9$ μs). The measurement coordinates on the wave front are denoted with the

TABLE I. Constant input parameters for numerical model.

Parameter	Value	Accuracy
c_f [m/s]	1487	±0.1%
c_p [m/s]	5958	±0.1%
c_s [m/s]	3268	±1%
ϱ^f [kg/m^3]	998	±1%
ϱ^s [kg/m^3]	7850	±1%

numbers 0,...,9, representing the subsequent angles of incidence (Fig. 1). This method was chosen because it offers a relatively simple opportunity to get an idea about the accuracy of the wave speeds (Table I). One of the results for the *P*-wave configuration coordinate 8° is presented in Fig. 7(a) and (b), for both the time domain and the frequency domain, respectively.

A typical measurement configuration during the *S*-wave measurements is depicted in Fig. 8. As illustrated by Fig. 8, the slab of steel was rotated over an angle of 10°. As in the case of the *P*-wave measurements, the incidence of a spherical wave front was simulated by rotating the transmitter over a subsequent set of angles of incidence. The measurement positions at the diffracted wave front ($T_{\mathrm{ar}}=407.9\ \mu$s) are indicated with the numbers 8,...,14, representing the angles of incidence θ_i in degrees (Fig. 8). In Fig. 9(a) and (b) the results for the shear wave are presented. As can be observed from Fig. 9(a) and (b), the comparison for the *S* wave between theory and experiment is quite good, but not as good as in case of the *P* wave. This is probably caused by the

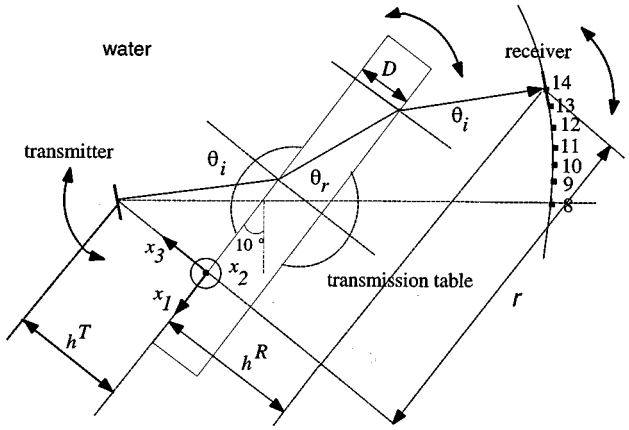


FIG. 8. Measurement configuration during *S*-wave measurements.

measuring accuracy for the *S*-wave speed which is a factor 10 lower than for the *P* wave. To show that the quasi plane-wave approximation is also valid in the particular transmission configuration under consideration we present the corresponding phase spectra in Fig. 10. From the presented phase spectrum of the *P* wave it is observed that the phase angle [cf. Eq. (112)] linearly relates to the frequency for phase angles between $[-\pi,\pi]$. This is typical for plane waves (see also Fig. 6).

B. The porous medium (fused glass beads)

The constituents of the porous medium are characterized by the parameters $\{\varrho^s,\kappa^s,\mu\}$ and $\{\varrho^f,\kappa^f\}$ for the solid and

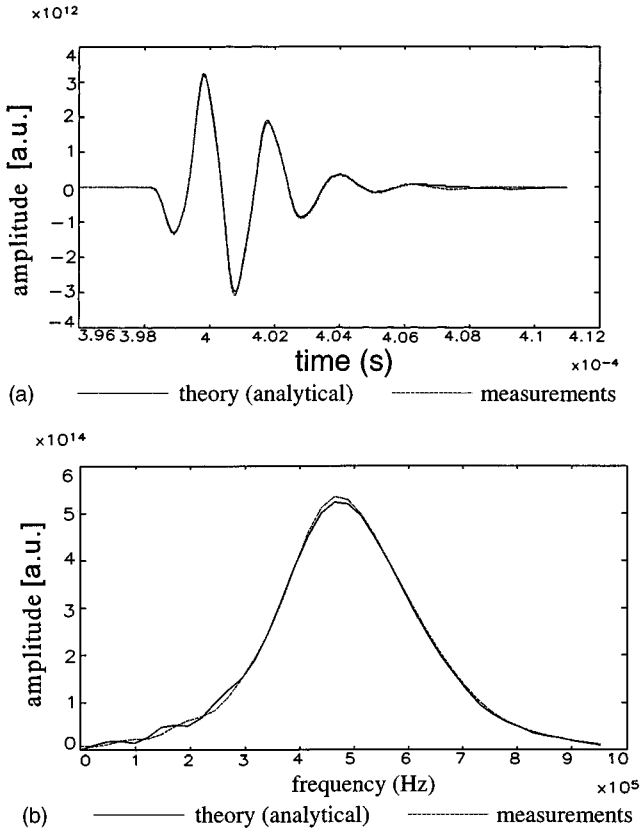


FIG. 7. (a) The *P* wave ($\theta_i=8°$), theory versus experiment (time domain). (b) The *P* wave ($\theta_i=8°$), theory versus experiment (frequency domain).

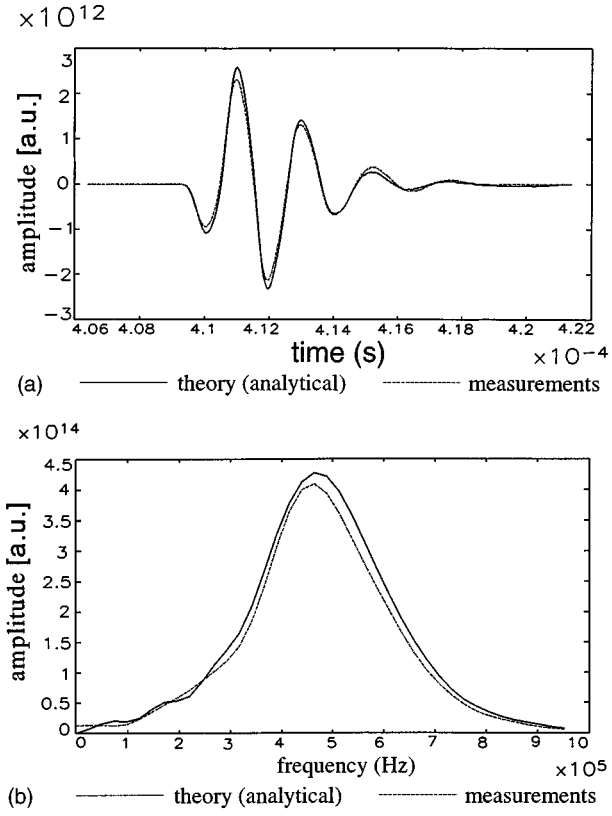


FIG. 9. (a) The *S* wave ($\theta_i=14°$) in the time domain: theory versus experiment. (b) The *S* wave ($\theta_i=14°$) in the frequency domain: theory versus experiment.

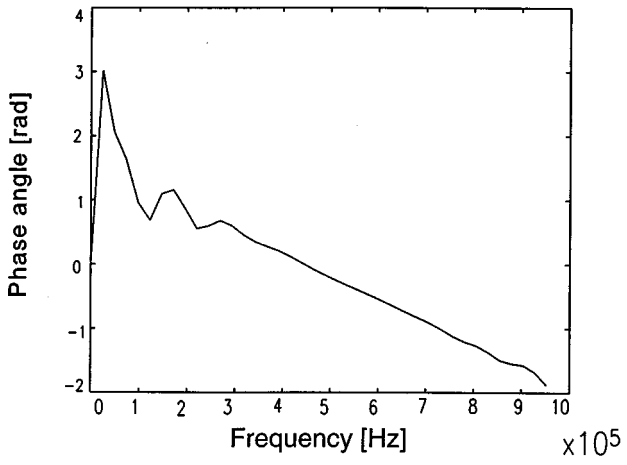


FIG. 10. Phase spectrum of P-wave ($\theta_i = 8°$).

the fluid phase, respectively. Again, the constitutive parameters of the solid phase follow from the elastodynamic wave speeds [Eqs. (114) and (115)]. The compressibility of the fluid phase follows from the fluid wave speed. The coupling between the fluid and solid phase in terms of the mutually induced mass density and constitutive parameters, respectively, is characterized by the parameters $\{m^{sf}, \kappa^{sf}, \mu^{ss}\}$, where we have assumed the porous medium to be reciprocal.[1] Once the mass densities and constitutive parameters of the constituting phases are determined, the coupling parameters, $\{m^{sf}, \kappa^{sf}, \mu^{ss}\}$ can be calculated from the porous medium wave speeds. In this procedure three fourth-order polynomials in terms of the coupling parameters $\{m^{sf}, \kappa^{sf}, \mu^{ss}\}$ follow from the wave speeds. The coefficients occurring in these polynomials are real and depend on the wave speeds of the porous medium and the mass densities and constitutive parameters of the constituting phases. From these three polynomials, four solutions for each of the coupling parameters $\{m^{sf}, \kappa^{sf}, \mu^{ss}\}$ follow. However, when we require the validity of the relations that make the volume densities of kinetic and deformation energy definite positive,[1] it appears that only one solution for each of the coupling parameters remains feasible. This procedure[2] resulted in unique values of the coupling coefficients $\{m^{sf}, \kappa^{sf}, \mu^{ss}\}$ for all glass bead fractions.

To verify de Vries' theory [cf. Eqs. (3)–(6)] in a transmission configuration, we used a porous skeleton of fused glass beads, saturated with water. The surrounding fluid in the water tank (upper and lower fluid half-space) is exactly the same as the saturating fluid in the porous medium. In Table II some physical constants of glass are listed.

The measurements that will be presented in this section relate to three types of porous media which only differ in

TABLE II. Physical constants for glass.

Parameter	Value	Source
ϱ^s [kg/m^3]	2200	Handb. of Phys.
c_p [m/s]	5900	Handb. of Phys.
c_s [m/s]	3750	Handb. of Phys.
κ^s [GPa]$^{-1}$	0.0283	Eq. (8.2.2)

TABLE III. Physical constants for fluid-saturated porous media.

Parameter	40–100 μm (0C)	160–250 μm (1C)	100–160 μm (2C)
c_+ [m/s]	3304	3294	3745
c_- [m/s]	956	952	913
c_S [m/s]	1764	1938	2138
λ_+ [mm]	6.6	6.6	7.5
λ_- [mm]	1.9	1.9	1.8
λ_S [mm]	3.5	3.9	4.3
λ_+/D_{dom} [−]	94	32	57
λ_-/D_{dom} [−]	27	9*	14*
λ_S/D_{dom} [−]	50	18	32
m^{sf} [kg/m^3]	−1096.5	−1069.2	−1375.5
κ^{sf} [GPa]$^{-1}$	−0.0353	−0.0499	−0.0225
μ^{ss} [GPa]	8.47	10.2	12.7
D [m]	0.0196	0.0195	0.0197
D/λ_+ [−]	3	3	2
D/λ_- [−]	10	10	10
D/λ_S [−]	5	5	4
ϱ^f [kg/m^3]	998	998	998
c_f [m/s]	1477	1476	1472
κ^f [GPa]$^{-1}$	0.459	0.46	0.462

dominant grain size (D_{dom}). In this way the permeability is the only quantity that varies, the porosity remaining constant. In Table III the physical constants of the fluid-saturated porous media are listed for the three different grain sizes. Further, λ_+, λ_-, and λ_s are evaluated at a reference frequency of $f = 0.5$ MHz, being the center frequency of the source signature (Fig. 5). The coefficients $\{m^{sf}, \kappa^{sf}, \mu^{ss}\}$ were calculated from the wave speeds ($c_\pm, c_S$) and the physical constants in Table II. With the information presented in Tables II and III, the Green's function $W^G(\mathbf{x}^R, t)$ is calculated. Subsequently the acoustic pressure at an arbitrary receiver coordinate in the lower fluid half-space is determined according to Eq. (113).

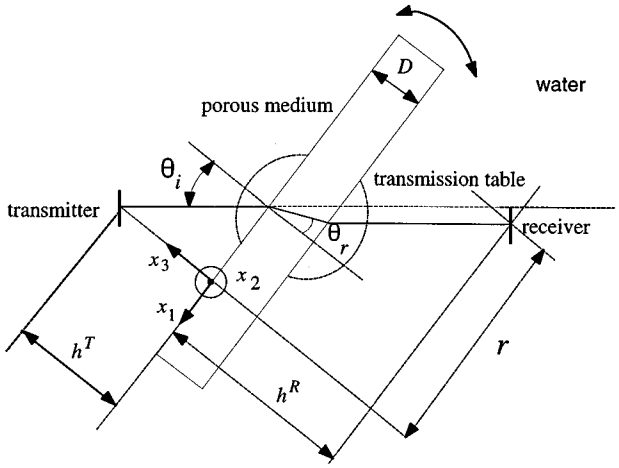


FIG. 11. Transmission configuration during porous media experiments.

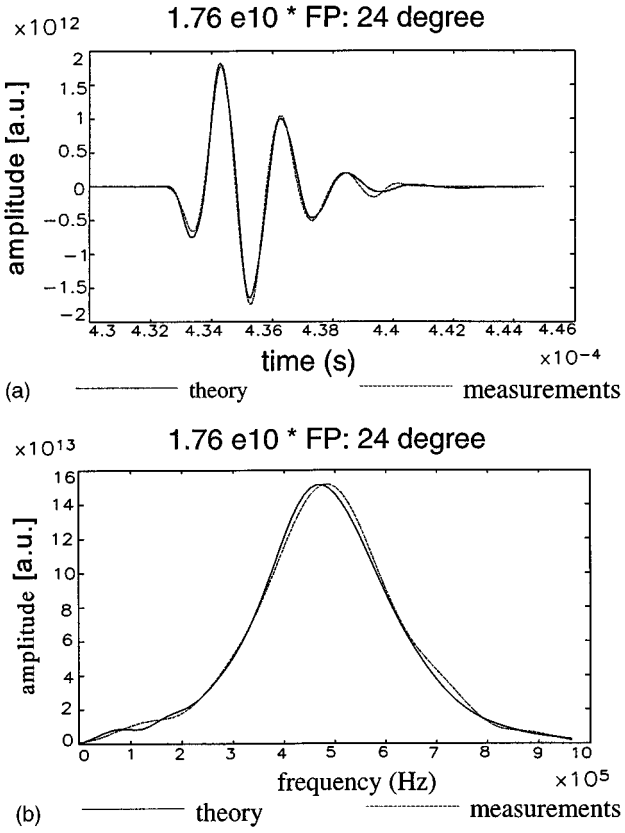


FIG. 12. (a) Theory versus experiment: FP, 0° (pressure versus time). Theory versus experiment: FP, 24° (pressure versus time). (b) Theory versus experiment: FP, 0° (amplitude spectrum). Theory versus experiment: FP, 24° (amplitude spectrum).

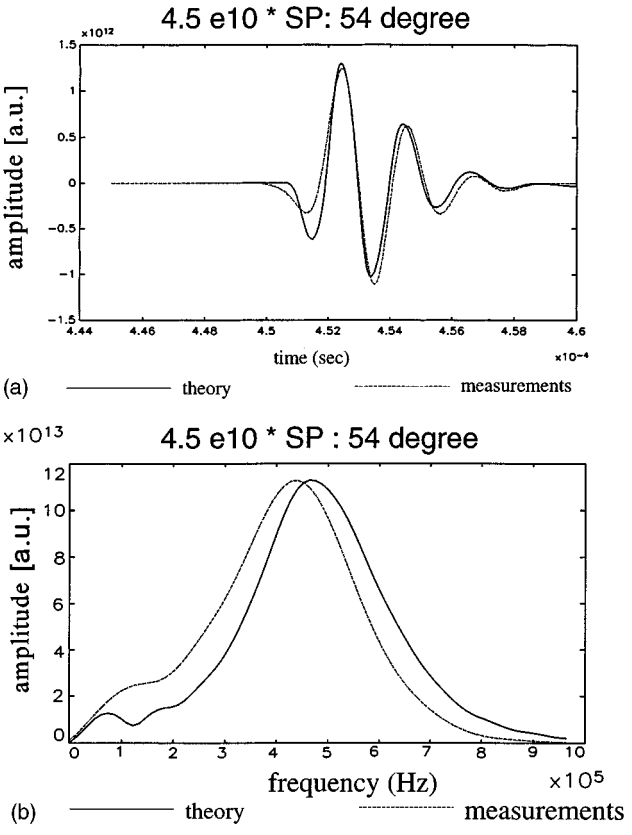


FIG. 13. (a) Theory versus experiment: SP, 54° (pressure versus time). (b) Theory versus experiment: SP, 54° (amplitude spectrum).

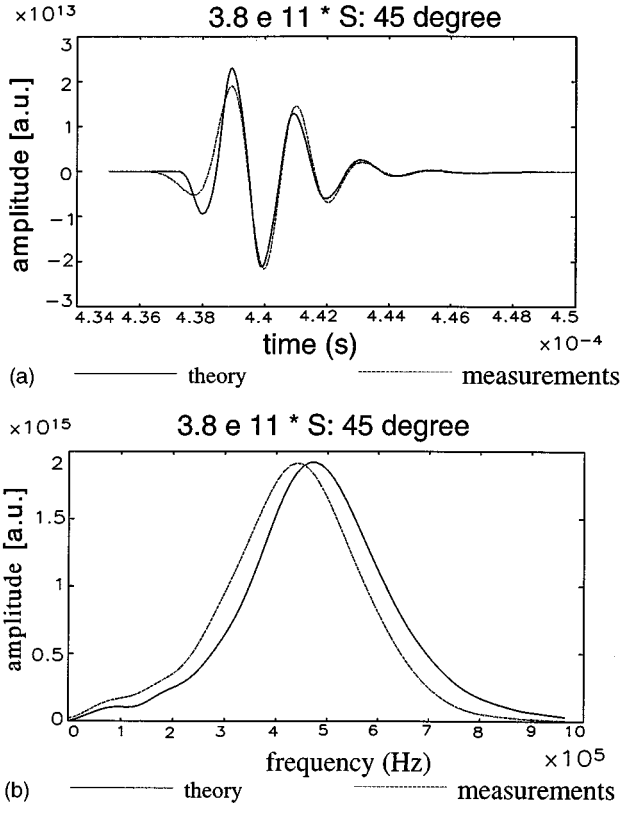


FIG. 14. (a) Theory versus experiment: S, 45° (pressure versus time). (b) Theory versus experiment: S, 45° (amplitude spectrum).

We will now present the measured as well as the corresponding synthetic seismograms for the fast and slow *P* waves and the shear wave for the grain size distribution: 40 $\leqslant D_{\mathrm{dom}} \leqslant 100$ μm. The comparison results for the other grain distributions will not be discussed in this paper because it followed from a microscopic analyses of corresponding thin sections that these grain fractions might be influenced by macroscopic inhomogeneities.[2] Furthermore, some of the measurements concerning these fractions are influenced by the phenomenon scattering, i.e., scattering by the pores. The bulk wave measurements which might be influenced by this phenomenon are denoted with an asterisk *. A typical transmission measurement setup for the fluid-saturated porous medium is presented in Fig. 11. In contrast to the measurements performed on the slab of steel, no wave front was scanned during the porous medium experiments. The angle of incidence was increased by rotating the slab of porous material on a transmission table, while the angle of incidence during the steel experiments was increased by rotating the transmitter.

Next, we will present the measured and synthetic seismograms. In Fig. 12(a) and (b) the results for a *fast P-wave* measurement at 24° are presented. Figure 12(a) shows the comparison between measured and synthetic seismograms in the time domain, and Fig. 12(b) the corresponding frequency domain representation. In Fig. 13(a) and (b) the measured and synthetic seismograms and corresponding frequency domain representations for a *slow P wave* at 54° are presented. Theoretically, all calculated slow *P* waves are influenced by

one head wave, being a critical fast P wave. This contribution relative to the body wave, however, is negligible.[2] Finally, Fig. 14(a) and (b) presents the results for a *shear wave* at 45°. Again, the head wave contributions are negligible with respect to the body wave contribution.[2]

C. Discussion

All presented figures of measured and synthetic seismograms are modified by a scale factor. This scale factor was calculated by dividing the theoretically calculated amplitude spectrum (solid lines in figures) by their corresponding measured amplitude spectrum at the center frequency. The measured seismograms were multiplied by this scale factor. From these scale factors and the presented figures it may be concluded that there is good agreement between theory and experiment, as far as the fast P wave is concerned. With respect to the slow P wave and the S wave, it is observed that the scale factors are different and that there is a frequency-dependent attenuation giving rise to a shift of the center frequency to a lower frequency. Furthermore, it is observed that the higher frequencies are more attenuated than the lower frequencies. The observed discrepancies are probably due to some kind of loss mechanism, which means that the lossless theory[1] is not an adequate description of acoustic wave propagation through the porous media investigated in this paper.

According to Biot,[4–6] there is a critical frequency, f_c, above which the inertia effects become predominant relative to the viscous loss effects. The critical frequency, f_c, is given by the equation

$$f_c = \frac{\eta \phi^f}{2\pi k \alpha_\infty \rho^f}, \tag{116}$$

where k is the Darcy permeability [m^2], ϕ^f is the porosity [$-$], α_∞ is the tortuosity, η is the fluid dynamic viscosity [Pa s], and ϱ^f is the fluid mass density [kg m]. Typical values for these constants with respect to the 0C fraction ($40 \leqslant D_{\mathrm{dom}} \leqslant 100\ \mu\mathrm{m}$) are:[2] $k = 1.8\times10^{-10}\ \mathrm{m}^2$, $\phi^f = 0.33$, $\alpha_\infty = 2.5$, $\eta = 1\times10^{-3}$ Pa s, $\varrho^f = 998\ \mathrm{kg\ m}^{-3}$. From these values it follows, according to Eq. (116), $f_c = 117$ Hz. Consequently, Biot predicts that viscous losses are very unlikely to be observable at frequencies of say $f > 0.2$ MHz (see Fig. 5), in the 0C fraction. Several researchers[8–11] compared measurements on these kind of artificial rock samples in this ultrasonic frequency range with the full frequency range Biot theory and showed the strong predictive power of the Biot theory and the attenuation mechanism involved. Kelder and Smeulders,[11] in fact, showed that even the frequency shift observed in our measurements is noticeable using the full frequency range Biot theory. However, a quantitative comparison between the full frequency range Biot theory and the measurements presented in this article has never been done so far. At this moment it is conjectured that the current measurements could be explained quantitatively by an extension of *De Vries*[1] theory in which inertial losses of the interface type are incorporated, as described in Ref. 3. These losses are assumed to be generated due to strong changes in pore geometry (tortuosity). Note, however, that the parameter, quantifying the losses (R), may depend on fluid viscosity, just as predicted by Johnson.[22] However, this extended theory still has to be verified.

VIII. CONCLUSION

A reliable technique based on the Fraunhofer diffraction theory was developed in order to compare laboratory transmission measurements with corresponding numerical simulations. The validity and accuracy of this technique was verified in a perfectly elastic medium transmission configuration and showed very good results.

Concerning the fluid-saturated porous medium investigated in this paper, it can be concluded that there is good agreement between theory and experiment as far as the fast compressional wave is concerned. The slow compressional wave and the shear wave, however, show evidence of some type of loss mechanism. It is very likely that the observed losses can be explained by incorporating Johnson's loss mechanism in the lossy counterpart of De Vries' theory.[3] It therefore seems profitable to verify both Biot's full frequency range theory[4–6] and the theory presented in Ref. 3, in the same transmission configuration presented in this paper.

ACKNOWLEDGMENTS

This work was supported by the European Community (E.C.). In particular the authors wish to thank Professor De Hoop, Professor Fokkema, Professor Peeters, Sijtze De Vries, and David Smeulders for the many fruitfull discussions about this subject. Furthermore, the authors thank Karel Heller for his technical support, Gerard Kwantes and Robert Pinke for their contributions during the experiments, and finally, Joop van Baaren for initiating this research project.

[1] S. M. de Vries, ''Propagation of transient acoustic waves in porous media,'' Ph.D. thesis, Delft University of Technology, 1989.
[2] T. W. Geerits, ''Acoustic wave propagation through porous: Theory and experiments,'' Ph.D. thesis, Delft University of Technology, 1993.
[3] T. W. Geerits, ''Acoustic wave propagation through porous media, revisited,'' J. Acoust. Soc. Am. **100**, 2949–2959 (1996).
[4] 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**, 168–178, 179–191 (1956).
[5] M. A. Biot, ''Generalized theory of acoustic propagation in porous dissipative media,'' J. Acoust. Soc. Am. **34**, 1254–1264 (1962).
[6] M. A. Biot, ''Mechanics of deformation and acoustic propagation in porous media,'' J. Appl. Phys. **33**, 1482–1498 (1962).
[7] T. J. Plona, ''Observation of a second bulk compressional wave in a porous medium at ultrasonic frequencies,'' Appl. Phys. Lett. **36**, 259–261 (1980).
[8] J. G. Berryman, ''Confirmation of Biot's theory,'' Appl. Phys. Lett. **87**, 382–384 (1980).
[9] D. L. Johnson, T. J. Plona, C. Scala, F. Pasierb, and H. Kojima, ''Tortuosity and acoustic slow waves,'' Phys. Rev. Lett. **49**, 1840–1844 (1982).
[10] D. L. Johnson, T. J. Plona, and H. Kojima, ''Probing porous media with first and second sound. II. Acoustic properties of water-saturated porous media,'' J. Appl. Phys. **76**, 115–125 (1994).
[11] O. Kelder and D. M. J. Smeulders, ''Measurement of ultrasonic bulk properties of water-saturated porous media,'' Extended Abstracts, 58th EAGE Mtg., C025 (1996).
[12] R. Burridge and J. B. Keller, ''Poroelasticity equations derived from microstructure,'' J. Acoust. Soc. Am. **70**, 1140–1146 (1981).
[13] S. R. Pride, A. F. Gangi, and F. D. Morgan, ''Deriving the equations of

motion for porous isotropic media,'' J. Acoust. Soc. Am. **6**, 3278–3290 (1992).
[14] H. Deresiewicz and R. Skalak, ''On uniqueness in dynamic poroelasticity,'' Bull. Seismol. Soc. Am. **53**, 783–788 (1963).
[15] L. E. Malvern, *Introduction to the Mechanics of a Continuous Medium*, edited by J. B. Reswick and W. M. Rohsenow (Prentice-Hall, Englewood Cliffs, NJ, 1969).
[16] A. T. de Hoop and J. H. M. T. van der Hijden, ''Generation of acoustic waves by an impulsive point source in a fluid/solid configuration with a plane boundary,'' J. Acoust. Soc. Am. **75**, 1709–1715 (1984).
[17] A. T. de Hoop, ''Acoustic radiation from impulsive sources in a layered fluid,'' Nieuwachief voor wiskunde **6**, 111–129 (1988).
[18] B. L. N. Kennett, ''Theoretical reflection seismograms for elastic media,'' Geophys. Prosp. **27**, 301–321 (1979).
[19] A. T. de Hoop, ''A modification of Cagniard's method for solving seismic pulse problems,'' Appl. Sci. Res. Sec. B **8**, 349–356 (1960).
[20] D. V. Widder, *The Laplace Transform* (Princeton U.P., Princeton, NJ, 1946).
[21] J. Krautkrämer, H. Krautkrämer, and W. Grabendorfer, *Ultrasonic Testing of Materials* (Springer-Verlag, Berlin, 1983), 3rd ed.
[22] 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).

Acoustic scattering by a hard or soft body across a wide frequency range by the Helmholtz integral equation method

Shih-An Yang
Department of Naval Architecture and Marine Engineering, National Cheng Kung University, Tainan, Taiwan

(Received 26 September 1996; revised 20 May 1997; accepted 2 July 1997)

In this paper, the acoustic scattering by an obstacle across a wide frequency range of sound waves is investigated on the basis of the Helmholtz integral formulation. To overcome the nonuniqueness difficulties, the methods proposed by Burton [NPL Report NAC 30 (Jan 1973)] and by Burton and Miller [Proc. R. Soc. London, Ser. A **323**, 201–210 (1971)] are adopted for the Dirichlet and Neumann problems, respectively. The aim of this paper is twofold. The first is to bring together completely regular formulations of the Helmholtz integral equation and its normal derivative. The second is to extend these formulations to treat the higher-frequency problems. The weakly singular integrals are regularized by subtracting out one term and adding it back. Depending on the problem concerned, the additional integral can finally be expressed in an explicit form or results in solving a surface source distribution of the equipotential body. The hypersingular kernels are regularized by the method of using some properties of the associated Laplace equation, originally proposed by Chien *et al.* [J. Acoust. Soc. Am. **88**, 918–937 (1990)]. The completely regularized integral equations are amenable to computation by direct use of the standard quadrature methods. To study the acoustic scattering due to higher-frequency waves, Filon's quadrature method [Proc. R. Soc. Edinburgh **49**, 38–47 (1928)] is extended to treat the rapidly oscillatory integrands. Numerical examples consist of acoustic scattering from a hard or soft sphere of radius a across a wide spectrum of wave numbers $ka=\pi-20\pi$. Comparisons of the numerical results with the exact solutions demonstrate the validity and efficiency of the implementation. [S0001-4966(97)04610-9]

PACS numbers: 43.20.Fn, 43.20.Tb [JEG]

INTRODUCTION

The surface Helmholtz integral equation is a common approach for the problems of the acoustic scattering by obstacles. The reason is that this approach has the advantages of reducing the dimensionality of the problem by one and transforming an infinite domain to finite boundaries in which the far-field radiation condition is satisfied automatically. It is well-known that the nonuniqueness problem occurs when using the Helmholtz integral equation alone. Two main approaches have been proposed to resolve the nuisance for the case of acoustically hard body scattering. The first method is called CHIEF, the combined Helmholtz integral equation formulation, due to Schenck,[1] which is perhaps the most widely used in the engineering applications. A series of works has been reported after Schenck, such as Refs. 2–7 to name some of them. The second method is due to Burton and Miller,[8] which combines the Helmholtz integral equation with its normal derivative. Comparison of these two methods shows that they introduce their own particular complications.[9–12]

The numerical difficulties of the Burton and Miller formulation arise mainly from the hypersingularity embedded in the integrodifferential equation. The tangential derivative formulation has been derived to regularize the highly singular kernels.[13,14] This approach alleviates the highly singular behavior only, and leaves an additional $1/R$ type of singularity in the integrand in the final formulation. To evaluate the lower-order singular integrals, special quadrature schemes are needed for the accuracy. The highly singular integral equations can also be interpreted as the Hadamard finite-part sense.[15,16] This implementation however is not straightforward and not popular in the practical applications. Chien *et al.*[17] employed certain identities for the hypersingular integrals arising in the associated integral equation for the Laplace equation in the interior domain. The order of singularity is reduced and further regularized after a polar coordinate transformation and some other special methods. Hwang[18] refined the process and obtained completely regular integral equations for the easier numerical computation.

The nonuniqueness also occurs in the case of the acoustic scattering by soft bodies. It was shown by Leis,[19] Panich,[20] and Brakhage and Werner[21] that the difficulty can be avoided by employing a suitable combination of single- and double-layer potentials. Also, the coupling factor must be chosen to be strictly complex when the wave frequency is real. Burton[22] later obtained similar formulation and showed that the integral operator of his formulation is the transpose of that appearing in the method of Brakhage and Werner.[21] The problem of acoustic scattering by soft bodies has been more or less overlooked for a long time. The reasons could be that the numerical implementation for the Cauchy principal value type of singularity is much easier compared with that for the hypersingularity, and that the soft scatterers are less useful in the engineering applications.

Integrals with rapidly oscillatory functions frequently occur in investigation in mathematical physics, and their

computation by quadratures is often desirable. From the previous experience of the boundary integral (element) method applied to acoustic problems, 5–6 nodal points within each wavelength are required for accurate numerical results. Smaller panels used to discretize the body surface are necessary when the wavelength decreases. Thus the dimension of the matrix to be inverted becomes greater if the Gauss elimination method is used. Tobocman[4,23] used the Padé approximation method in place of matrix inversion by Gauss–Jordon elimination to reduce the need for large amounts of random access memory and computing time in shorter wavelength cases. The scattering of acoustic waves by a rigid sphere of unit radius was studied for the values of wave number varied from 3 to 21. Filon[24] computed integrals of the form $\int_a^b f(x)\exp(ikx)dx$, where i is the imaginary unit. The interval of integration was divided into several parts and $f(x)$ of each small interval was approximated by parabolic arcs. For parabolic $f(x)$, the integrals can be computed explicitly by integration by parts. Luke[25] extended this method by approximating $f(x)$ by piecewise polynomials of higher degree. A generalization of Filon's method due to Smith,[26] who considered a nonlinear argument in the oscillating functions, was used by Shen and Farell[27] for the prediction of wave resistance for a moving ship.

In this article we attempt to develop an efficient numerical method that is applicable for the scattering problems across a wide frequency range. The completely singularity-free formulations of the Helmholtz integral equation and its normal derivative will be outlined in the first phase. Those formulations are preferable and of importance in terms of the numerical point of view. The present work will try to transform the weakly singular integrals to regular ones, in which the surface source distribution of the corresponding equipotential body of the same geometry has to be calculated. The regularization method for hypersingular kernels, derived by Chien *et al.*[17] and later by Hwang,[18] will be summarized. All the resultant integrals are regular and therefore are amenable to direct use of the standard quadrature methods. The mathematical derivation of the abovementioned regularization procedures will be described in Secs. II and III after a brief introduction of the Helmholtz integral equation in Sec. I. An extension of Filon's method, which is suitable for the acoustic scattering problems with short wavelengths, will be described in Sec. IV. Section V demonstrates the numerical results of the scattering by a sphere of radius a subject to Dirichlet and Neumann boundary conditions, respectively. The values of wave number ka will be tested from π to 20π. Comparisons of the numerical results with the known exact solutions will be made and show the validity and efficiency of the implementation. Section VI contains the conclusions.

I. A BRIEF SUMMARY OF THE INTEGRAL EQUATION FORMULATION

The integral formulation of the associated acoustic problems is well-known and can be found in books or literature. For the sake of completeness and the introduction of notation, we summarize the relevant fundamentals in this section. The governing equation for the propagation of acoustic waves through an unbounded homogeneous medium is described by the wave equation

$$\nabla^2\phi(r,t)=\frac{1}{c^2}\frac{\partial^2\phi(r,t)}{\partial t^2}, \tag{1}$$

where ∇^2 denotes the Laplacian operator in three dimensions, ϕ the velocity potential at a point r at time t, and c the speed of sound in the medium at the equilibrium state. The velocity potential ϕ relates to the velocity by

$$\mathbf{u}=\nabla\phi, \tag{2}$$

and can be expressed as

$$\phi=\phi^i+\phi^s, \tag{3}$$

where ϕ^i is the incident velocity potential and ϕ^s is the scattered velocity potential. The excess acoustic pressure can be written as

$$p=-\rho_0\frac{\partial\phi}{\partial t}, \tag{4}$$

where ρ_0 is the density of the fluid at the equilibrium state. The acoustic pressure can also be represented as a sum of two parts

$$p=p^i+p^s, \tag{5}$$

where p^i is the incident pressure and p^s is the scattered pressure. The Helmholtz differential equation is obtained from Eq. (1) for time-harmonic waves with a time factor $\exp(-i\omega t)$, i.e.,

$$(\nabla^2+k^2)\phi=0, \tag{6}$$

where i is the imaginary unit, ω the angular frequency, and $k=\omega/c$ the wave number. Equation (4) becomes

$$p=i\omega\rho_0\phi. \tag{7}$$

At the surface of a soft scatterer, the excess pressure is zero; hence

$$\phi=0. \tag{8}$$

At the surface of a hard scatterer, the normal component $\mathbf{u}\cdot\mathbf{n}$ of the particle velocity $\mathbf{u}$ is zero; hence

$$\frac{\partial\phi}{\partial n}=0, \tag{9}$$

where $\mathbf{n}$ is the unit vector normal to the surface of the body and into the surrounding space, and n is the distance along the external normal vector $\mathbf{n}$. Conditions (8) and (9) are known as Dirichlet and Neumann boundary conditions, respectively. The scattered velocity potential should also satisfy the Sommerfeld radiation condition, which is expressed in three dimensions as

$$\lim_{r\to\infty} r\left|\frac{\partial\phi^s}{\partial r}-ik\phi^s\right|=0. \tag{10}$$

The equivalent boundary integral formulation of Eq. (6), which is valid for an acoustic medium B' exterior to a finite body B with surface S takes the following form:

$$\epsilon\phi(P)=\phi^i(P)+\int_S\left(\phi(Q)\,\frac{\partial G_k(P,Q)}{\partial n_Q}-G_k(P,Q)\,\frac{\partial\phi(Q)}{\partial n_Q}\right)dS_Q. \tag{11}$$

The radiation boundary condition has been taken into account in Eq. (11). The free-space Green's function G_k for the Helmholtz equation in three dimensions is given by

$$G_k(P,Q)=e^{ikR}/4\pi R, \tag{12}$$

where R is the distance between the field point P and the moving point Q, and n_Q is the outward directed normal at Q. The value of factor ϵ depends on the position of P, i.e.,

$$\begin{aligned}\epsilon&=0, \quad \text{for } P\in B,\\ &=1, \quad \text{for } P\in B,\\ &=1/2, \quad \text{for } P\in \text{smooth } S.\end{aligned}$$

For P on the surface with a sharp corner, ϵ relates to the solid angle α by

$$\begin{aligned}\epsilon&=1-\alpha/4\pi\\ &=1+\frac{1}{4\pi}\int_S\frac{\partial}{\partial n_Q}\left(\frac{1}{R(P,Q)}\right)dS_Q.\end{aligned} \tag{13}$$

The integral equation (11) becomes singular at certain critical frequencies. The nonuniqueness of the solution is a purely mathematical problem, i.e., there is no physical meaning connecting to it. In order to overcome the difficulty of nonuniqueness, a second integral equation can be formed by differentiating Eq. (11) in the normal direction at P:

$$\epsilon\frac{\partial\phi(P)}{\partial n_P}=\frac{\partial\phi^i(P)}{\partial n_P}+\frac{\partial}{\partial n_P}\int_S\phi(Q)\,\frac{\partial G_k(P,Q)}{\partial n_Q}\,dS_Q-\int_S\frac{\partial G_k(P,Q)}{\partial n_P}\,\frac{\partial\phi(Q)}{\partial n_Q}\,dS_Q. \tag{14}$$

By a linear combination of the two, unique solutions can be obtained for all real frequencies when the coupling parameter is chosen such that its imaginary part is nonzero.[22,8]

II. REGULARIZATION OF SINGULAR KERNELS FOR THE SOFT BODY SCATTERING

For a soft body of smooth geometry, Eq. (11) reduces to

$$0=\phi^i(P)-\int_S G_k\,\frac{\partial\phi(Q)}{\partial n_Q}\,dS_Q. \tag{15}$$

This equation involves singularities when the moving point coincides with the fixed point. There are numerical techniques available for this $1/R$ type of singularity (e.g., Lachat and Watson,[28] Seybert and co-workers[5,7]). A new formulation however will be developed to take care of the singular kernels. The idea, which is a classic one in the fields of potential theory (e.g., Landweber and Macagno[29]) and acoustics, is to subtract a function from the integrand so that the resultant kernel becomes regular, and then add an accurate integration of the function back to the equation. Thus Eq. (15) is rearranged as

$$0=\phi^i(P)-\int_S(G_k-G_0)\,\frac{\partial\phi(Q)}{\partial n_Q}\,dS_Q-\int_S G_0\left(\frac{\partial\phi(Q)}{\partial n_Q}-\frac{\partial\phi(P)}{\partial n_P}\right)dS_Q-\frac{\partial\phi(P)}{\partial n_P}\int_S G_0\,dS_Q, \tag{16}$$

where G_0 is the free-space Green's function for Laplace equation and can be written in three dimensions as

$$G_0=1/4\pi R. \tag{17}$$

We note that the first integral is regular, since the kernel is bounded as R tends to zero, i.e.,

$$\lim_{R\to0}(G_k-G_0)=\frac{ik}{4\pi}. \tag{18}$$

We now show that the second integral is also regular. When $Q(\xi,\eta,\zeta)$ approaches $P(x,y,z)$, we can write the Taylor expansion of the integrand in the second integral in the form:

$$\begin{aligned}&G_0\left(\frac{\partial\phi(Q)}{\partial n_Q}-\frac{\partial\phi(P)}{\partial n_P}\right)\\ &\quad=\frac{1}{4\pi R}\left\{\mathbf{R}\cdot\nabla_\xi\left(\frac{\partial\phi(P)}{\partial n_P}\right)+\frac{1}{2!}\left[(\xi-x)^2\left(\frac{\partial\phi(P)}{\partial n_P}\right)_{\xi\xi}+\cdots\right]+\cdots\right\}\approx O(R^0),\end{aligned} \tag{19}$$

where $\mathbf{R}=(\xi-x,\eta-y,\zeta-z)$. We observe that the leading term is bounded and antisymmetric, although its value is indeterminate, as R approaches zero. The integral of this integrand over a small area symmetric about P, however, is zero. For this reason we propose to set the integrand of the second integral of Eq. (16) equal to zero when Q coincides with P. The last integral is still singular in its present form but can be regularized in terms of the potential theory. Applying the property of equipotential body, we obtain

$$\begin{aligned}\int_S G_0\,dS_Q&=\int_S\left[1-\frac{\sigma(Q)}{\sigma(P)}\right]G_0\,dS_Q+\frac{1}{\sigma(P)}\int_S\sigma(Q)G_0\,dS_Q\\ &=\int_S\left[1-\frac{\sigma(Q)}{\sigma(P)}\right]G_0\,dS_Q-\frac{\Phi_e}{\sigma(P)},\end{aligned} \tag{20}$$

where σ is a source distribution on S and makes the surface an equipotential of potential Φ_e. By the definition,

$$\Phi_e=-\int_S\sigma(Q)G_0\,dS_Q. \tag{21}$$

We note first that the integrand on the right-hand side of Eq. (20) can be set to zero by using the Taylor expansion and then by the similar reasoning just stated above. Since the potential is constant in the interior of an equipotential surface, its value may conveniently be computed at the origin as

$$\Phi_e=-\frac{1}{4\pi}\int_S\frac{\sigma(Q)}{(\xi^2+\eta^2+\zeta^2)^{1/2}}\,dS_Q. \tag{22}$$

Since the velocity in the interior domain of an equipotential body vanishes, the source intensity σ can be written as

$$\sigma(P) = -2\int_S \sigma(Q)\,\frac{\partial G_0}{\partial n_P}\,dS_Q. \tag{23}$$

The singularity in Eq. (23) can be removed in the following form:

$$\sigma(P) = -2\int_S \left[\sigma(Q)\,\frac{\partial G_0}{\partial n_P} - \sigma(P)\,\frac{\partial G_0}{\partial n_Q}\right] dS_Q - 2\sigma(P)\int_S \frac{\partial G_0}{\partial n_Q}\,dS_Q. \tag{24}$$

Gauss's flux theorem states that at points of S where the normal is continuous

$$-\int_S \frac{\partial G_0}{\partial n_Q}\,dS_Q = \frac{1}{2}. \tag{25}$$

This theorem enables us to simplify Eq. (24) to

$$0 = \int_S \left[\sigma(Q)\,\frac{\partial G_0}{\partial n_P} - \sigma(P)\,\frac{\partial G_0}{\partial n_Q}\right] dS_Q, \tag{26}$$

which is a Fredholm integral equation of the first kind. We shall again show that the singularity in Eq. (26) when Q coincides with P is not present. Let $f(x,y,z)=0$ be the equation of the body surface S. The direction cosines $\mathbf{c}=(l,m,n)$ at the point P of S are given by

$$l_P = \frac{f_x}{D_P}, \quad m_P = \frac{f_y}{D_P}, \quad n_P = \frac{f_z}{D_P}, \tag{27}$$

or

$$\mathbf{c} = \frac{\boldsymbol{\nabla}_x f}{D_P}, \tag{28}$$

where f_x, f_y, f_z denote partial derivatives of $f(x,y,z)$ with respect to x, y, z, and

$$D_P = (f_x^2 + f_y^2 + f_z^2)^{1/2}. \tag{29}$$

Then we have

$$\frac{\partial G_0}{\partial n_P} = \frac{1}{4\pi R^3}\,\mathbf{R}\cdot\mathbf{c} = \frac{1}{4\pi R^3 D_P}\,\mathbf{R}\cdot\boldsymbol{\nabla}_x f, \tag{30}$$

so that

$$\sigma(Q)\,\frac{\partial G_0}{\partial n_P} - \sigma(P)\,\frac{\partial G_0}{\partial n_Q} = \frac{1}{4\pi R^3}\left[\frac{\sigma(Q)}{D_P}\,\mathbf{R}\cdot\boldsymbol{\nabla}_x f + \frac{\sigma(P)}{D_Q}\,\mathbf{R}\cdot\boldsymbol{\nabla}_\xi f\right]. \tag{31}$$

When Q is near P, we have the Taylor expansion

$$f(x,y,z) = f(\xi,\eta,\zeta) - \mathbf{R}\cdot\nabla_\xi f + \frac{1}{2!}\left[(x-\xi)^2 f_{\xi\xi} + \cdots\right] + \cdots. \tag{32}$$

Since $f(x,y,z)=f(\xi,\eta,\zeta)=0$, we obtain

$$\mathbf{R}\cdot\nabla_\xi f = \tfrac{1}{2}\left[(\xi-x)^2 f_{\xi\xi} + \cdots\right] + O(R^3). \tag{33}$$

Similarly we have

$$\mathbf{R}\cdot\boldsymbol{\nabla}_x f = \frac{-1}{2}\left[(\xi-x)^2 f_{\xi\xi} + \cdots\right] + O(R^3). \tag{34}$$

Also we have

$$\sigma(Q) = \sigma(P) + \mathbf{R}\cdot\nabla_x\sigma + \cdots \tag{35}$$

and

$$\frac{1}{D_Q} = \frac{1}{D_P} + \mathbf{R}\cdot\nabla_x\left(\frac{1}{D_P}\right) + \cdots. \tag{36}$$

Substituting Eqs. (33)–(36) into (31), we observe that terms of the first and second order cancel, leaving antisymmetric terms of the third order in the numerator. Since the denominator is also of the third order, the ratio is indeterminate as R approaches zero. We can then set the integrand in Eq. (26) equal to zero by the same reasoning as mentioned before. In the sequel, we obtain the singularity-free form of the Helmholtz integral equation (15) in the following:

$$0 = \phi^i(P) - \int_S (G_k - G_0)\,\frac{\partial\phi(Q)}{\partial n_Q}\,dS_Q - \int_S G_0\left(\frac{\partial\phi(Q)}{\partial n_Q} - \frac{\partial\phi(P)}{\partial n_P}\right) dS_Q - \frac{\partial\phi(P)}{\partial n_P}\times\int_S\left[1 - \frac{\sigma(Q)}{\sigma(P)}\right] G_0\,dS_Q + \frac{\partial\phi(P)}{\partial n_P}\,\frac{\Phi_e}{\sigma(P)}, \tag{37}$$

where the source intensity σ is given by Eq. (26) and the equipotential Φ_e is given by Eq. (22).

The normal derivative of the Helmholtz integral equation for the soft bodies takes the form

$$\frac{1}{2}\,\frac{\partial\phi(P)}{\partial n_P} = \frac{\partial\phi^i(P)}{\partial n_P} - \int_S \frac{\partial G_k}{\partial n_P}\,\frac{\partial\phi(Q)}{\partial n_Q}\,dS_Q. \tag{38}$$

Applying the same technique as in the previous case, we have

$$\frac{1}{2}\,\frac{\partial\phi(P)}{\partial n_P} = \frac{\partial\phi^i(P)}{\partial n_P} - \int_S\left[\frac{\partial G_k}{\partial n_P}\,\frac{\partial\phi(Q)}{\partial n_Q} - \frac{\partial G_0}{\partial n_Q}\,\frac{\partial\phi(P)}{\partial n_P}\right]\times ds_Q - \frac{\partial\phi(P)}{\partial n_P}\int_S \frac{\partial G_0}{\partial n_Q}\,dS_Q. \tag{39}$$

Using Gauss's flux theorem, we have

$$0 = \frac{\partial\phi^i(P)}{\partial n_P} - \int_S\left[\frac{\partial G_k}{\partial n_P}\,\frac{\partial\phi(Q)}{\partial n_Q} - \frac{\partial G_0}{\partial n_Q}\,\frac{\partial\phi(P)}{\partial n_P}\right] dS_Q, \tag{40}$$

where the integrand is set equal to zero as Q coincides with P.

We have now derived two singularity-free formulations of the Helmholtz integral equation and its normal derivative in Eqs. (37) and (40), respectively. We can then obtain a unique solution by a combination of these two equations with a coupling constant of nonzero imaginary part according to Burton's method.[22] It is clear that the direct quadrature methods can be applied to these formulations immediately. The additional expenditure of computing time of the proposed method is mainly due to the calculation of the source

distribution σ which creates the surface of the scatterer an equipotential. However, the solution of source intensity σ requires just a small amount of computing time in comparison with the solution of the acoustic field itself. The reason is that σ is a function of the geometry of the scatterer only. Thus σ is independent of the acoustic wave frequencies and the boundary conditions at the surface of the scatterer. The change of acoustic wave frequencies will not affect the intensity of σ. Therefore a set of collocation points which produces numerically an accurate σ can be irrelevant to that of collocation points for the acoustic field if we require so. The cases take place especially when the higher-frequency problems are encountered in which the adoption of higher-order quadrature methods is imperative for the convergence and accurate solutions. In these cases we can estimate the source intensity σ accurately at the collocation points of the denser system for the acoustic scattering by a proper interpolation method from the looser grid system for σ.

III. REGULARIZATION OF SINGULAR KERNELS FOR THE HARD BODY SCATTERING

For a hard body of smooth geometry, Eq. (11) reduces to

$$\frac{\phi(P)}{2}=\phi^i(P)+\int_S \phi(Q)\,\frac{\partial G_k}{\partial n_Q}\,dS_Q. \tag{41}$$

We can regularize Eq. (41) using the similar regularization procedures as in the last section; thus

$$\frac{\phi(P)}{2}=\phi^i(P)+\int_S\left[\phi(Q)\,\frac{\partial G_k}{\partial n_Q}-\phi(P)\,\frac{\partial G_0}{\partial n_Q}\right]dS_Q + \phi(P)\int_S \frac{\partial G_0}{\partial n_Q}\,dS_Q. \tag{42}$$

Applying Gauss's flux theorem, we obtain the Fredholm integral equation of the second kind

$$\phi(P)=\phi^i(P)+\int_S\left[\phi(Q)\,\frac{\partial G_k}{\partial n_Q}-\phi(P)\,\frac{\partial G_0}{\partial n_Q}\right]dS_Q, \tag{43}$$

where the integral is regular because the integrand is set to zero as Q coincides with P by the reasoning mentioned in the last section.

Taking the normal derivative of Eq. (41) at point P, we have

$$\frac{\partial \phi^i(P)}{\partial n_P}=-\frac{\partial}{\partial n_P}\int_S \phi(Q)\,\frac{\partial G_k}{\partial n_Q}\,dS_Q, \tag{44}$$

which contains a hypersingular kernel as point Q coincides with point P. The derivative $\partial/\partial n_P$ cannot be taken inside the integral since a hypersingular kernel is nonintegrable. Equation (44) is actually an integrodifferential equation. The hypersingular integral has been regularized by Chien *et al.*[17] and Hwang.[18] We shall not repeat the derivation and just outline the resultant singularity-free formulation in the following:

$$\frac{k^2\Phi_e}{2\sigma(P)}\,\phi(P)=\frac{\partial \phi^i(P)}{\partial n_P}-\frac{\partial \varphi(P)}{\partial n_P}+\int_S\left[\phi(Q)\,\frac{\partial^2(G-G_0)}{\partial n_P\,\partial n_Q} - \frac{k^2 G_0\sigma(Q)}{2\sigma(P)}\,\phi(P)\right]dS_Q + \int_S\left[\frac{\partial G_0}{\partial n_P}\,\frac{\partial \varphi(Q)}{\partial n_Q}-\frac{\partial G_0}{\partial n_Q}\,\frac{\partial \varphi(P)}{\partial n_P}\right]dS_Q, \tag{45}$$

where the source density σ and the equipotential Φ_e were defined in the last section. The solution of the additional variable $\partial\varphi/\partial n$ requires an additional equation,

$$\frac{\Phi_e}{\sigma(P)}\,\frac{\partial \varphi(P)}{\partial n_P}=\int_S G_0\left[\frac{\partial \varphi(Q)}{\partial n_Q}-\frac{\sigma(Q)}{\sigma(P)}\,\frac{\partial \varphi(P)}{\partial n_P}\right]dS_Q - \int_S[\phi(Q)-\phi(P)]\,\frac{\partial G_0}{\partial n_Q}\,dS_Q, \tag{46}$$

which is a result by applying an associated integral equation for the Laplace equation in the interior domain. Both of the integrands in Eq. (46) are set equal to zero as Q is close to P. The first and second integrands on Eq. (45) are set equal to $ik^3\phi(P)/12\pi$ and 0, respectively.

Similar to the problem for the soft body scattering, a combination of the surface Helmholtz equation (43) and its normal derivative Eq. (45) with a coupling factor which has nonzero imaginary part for the real wave number k will guarantee a unique solution. Up to this stage, we have completed the presentation of the completely regular formulations of the Helmholtz integral equation and its normal derivative subject to Dirichlet and Neumann boundary conditions, respectively. All of these formulations are amenable to the direct use of the conventional quadrature methods, if we consider low to mediate oncoming wave frequencies only. For the higher-frequency problem, a specific method is needed and will be presented in the next section.

IV. FILON'S QUADRATURE METHOD

We have obtained integral equations for solving the scattering of acoustic waves by an obstacle. The surface integrals involving the wave number k can be cast in the form

$$I=\int_S f e^{ikR}\,dS, \tag{47}$$

where f is a nonlinear function and R is the distance between the fixed point and the source point. For shorter wavelengths, the integrands will be highly oscillating so that the conventional quadrature methods become inefficient. Filon[24] worked out a method for the numerical implementation of the integrals $\int_a^b f(x)e^{ikx}\,dx$. In Filon's method, $f(x)$ is approximated in each integration subinterval by a parabola obtained by interpolation to $f(x)$ at the equally spaced node points. For parabolic $f(x)$, the integrals can be computed explicitly by integration by parts. Smith[26] extended Filon's method to the integrals of the type $\int_a^b f(x)e^{ig(x)}\,dx$, where $g(x)$ is a nonlinear quantity. The essential feature of Smith method is the replacement of the nonlinear function $g(x)$ over each integration subinterval by a linear function plus an increment $\delta(x)$, i.e.,

$$g(x)=\lambda x+\delta(x), \tag{48}$$

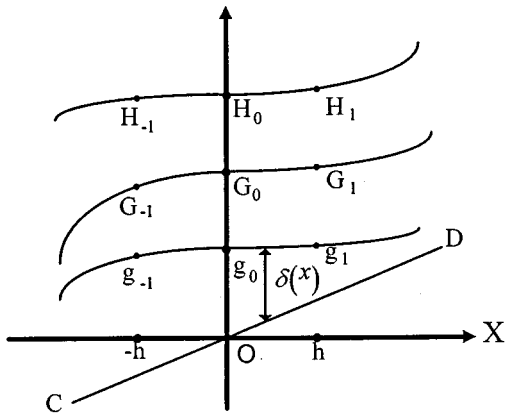


FIG. 1. Illustration of three-point quadrature treatment.

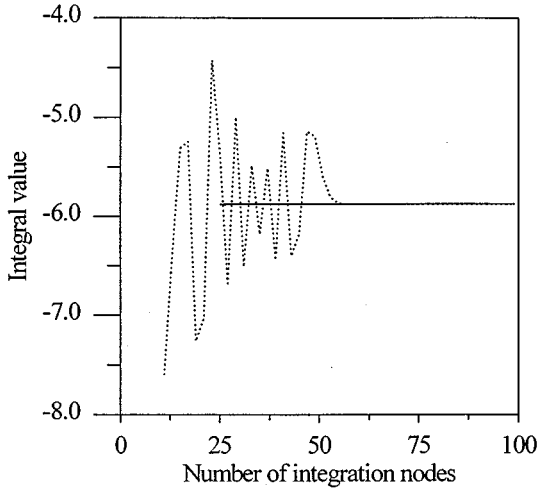


FIG. 2. Comparison between the present quadrature method (———) and the Gauss–Legendre method (····).

where $\delta(x)$ is the difference between $g(x)$ and the line CD, and λ is the slope of the line CD. Figure 1 illustrates the three-point quadrature treatment. Thus we have

$$\int_{-h}^{h} f(x)e^{ig(x)}dx = \int_{-h}^{h}[G(x)\cos \lambda x - H(x)\sin \lambda x]dx + i\int_{-h}^{h}[G(x)\sin \lambda x + H(x)\cos \lambda x]dx \equiv I_c + iI_s, \tag{49}$$

where $G(x)=f(x)\cos \delta(x)$ and $H(x)=f(x)\sin \delta(x)$. I_c and I_s can be integrated explicitly by integration by parts since $G(x)$ and $H(x)$ are approximated by a parabola at points $x=-h$, 0, and h. The explicit formulas for I_c and I_s are

$$I_c = h[2(k_0+k_2)G_0 - k_2(G_1+G_{-1}) + k_1(H_1-H_{-1})], \tag{50}$$

$$I_s = h[-k_1(G_1-G_{-1}) + 2(k_0+k_2)H_0 - k_2(H_1+H_{-1})], \tag{51}$$

where

$$k_0 = \frac{\sin \Theta}{\Theta}, \quad k_1 = \frac{-1}{\Theta}(k_0 - \cos \Theta), \quad k_2 = \frac{-1}{\Theta}(2k_1 + \sin \Theta), \tag{52}$$

and $\Theta=(g_1-g_{-1})/2$. It should be noted that roundoff errors will be a source of trouble when Θ is small. For this case the functions can be approximated as

$$k_2 = -\frac{1}{3} + \frac{\Theta^2}{5\cdot 2!} - \frac{\Theta^4}{7\cdot 4!} + \frac{\Theta^6}{9\cdot 6!} - \frac{\Theta^8}{11\cdot 8!} + \frac{\Theta^{10}}{13\cdot 10!} - \cdots, \tag{53}$$

$$k_1 = -\frac{1}{2}(\Theta k_2 + \sin \Theta), \quad k_0 = -(\Theta k_1 - \cos \Theta).$$

The two-dimensional version of Smith's formulas has been derived in a similar way. We shall not outline those lengthy formulas here. To evaluate the validity of the method, let us consider a single integral

$$I_{ex} = \int_0^2 \left(\frac{\cos kx}{x} - \frac{1}{x}\right)dx. \tag{54}$$

To accommodate the example integral to the present method, Eq. (54) is rewritten as

$$I_{ex} = -2\int_0^{\epsilon}\left(\frac{1}{x}\sin \frac{kx}{2}\right)\sin \frac{kx}{2}\,dx + \int_{\epsilon}^{2}\frac{1}{x}\cos kx\,dx - \int_{\epsilon}^{2}\frac{1}{x}\,dx, \tag{55}$$

where we set $\epsilon=0.07\pi$ and $k=100$. Figure 2 shows that the present method is superior to the Gauss–Legendre method. It is noted in passing that higher-order formulas can be derived in a similar way, although we applied three-point formulas in this paper.

V. NUMERICAL CALCULATIONS

The numerical examples are the problems of acoustic scattering from a soft or hard sphere with radius a. The origin of the spherical polar coordinates (r,θ^*,α^*) is located at the center of the sphere, where $0\leqslant\theta^*\leqslant\pi$ and $0\leqslant\alpha^*\leqslant 2\pi$. For a plane wave, $\phi^i=\exp(-ikz)$, incident in the direction of the negative z axis, $z=r\cos\theta^*$, the unknown variables were solved along the cone angle θ^*, and $\alpha^*=0$. The coupling factor was set to be i/ka. First, we would like to test the validity of the formulations presented in Secs. II and III. In this case the Gauss–Legendre method was used in the numerical calculations. On the surface of a soft sphere, the exact solution is given by

$$\frac{1}{k}\frac{\partial \phi}{\partial r} = -\frac{i}{(ka)^2}\sum_{n=0}^{\infty}(-i)^n(2n+1)\frac{P_n(\cos\theta^*)}{h_n^{(1)}(ka)}, \tag{56}$$

where P_n is the Legendre function of the first kind and $h_n^{(1)}$ is the spherical Hankel function. The associated exact solution for a hard sphere is given by

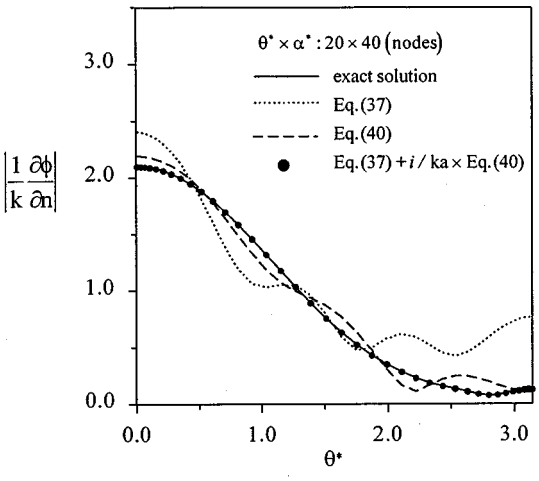


FIG. 3. Surface field for a soft sphere exposed to a plane wave with $ka = 4.4934$.

$$\phi = \frac{i}{(ka)^2} \sum_{n=0}^{\infty} (-i)^n (2n+1) \frac{P_n(\cos\theta^*)}{h_n^{(1)'}(ka)}. \tag{57}$$

The nondimensional solutions on the surface of the sphere as a function of θ^* are depicted in Figs. 3 and 4 for the Dirichlet and Neumann boundary conditions, respectively. Specifically, we set $ka = 4.4934$, which is a characteristic wave number for both the Helmholtz integral equation and its normal derivative. We observe that the two equations fail simultaneously in Figs. 3 and 4. A combination of the two equations gives fairly satisfactory results in comparison with the exact solutions. Thus the validity of the proposed formulations has been justified successfully.

The higher-frequency problem will be studied in the following. We assumed a constant variable distributed at each equally spaced segment along θ^*. Figure 5 compares the results by the present method and the conventional Gauss–Legendre quadrature method with the exact solutions for the case of soft sphere scattering. Both numerical methods applied the LU decomposition scheme for solving the matrices.

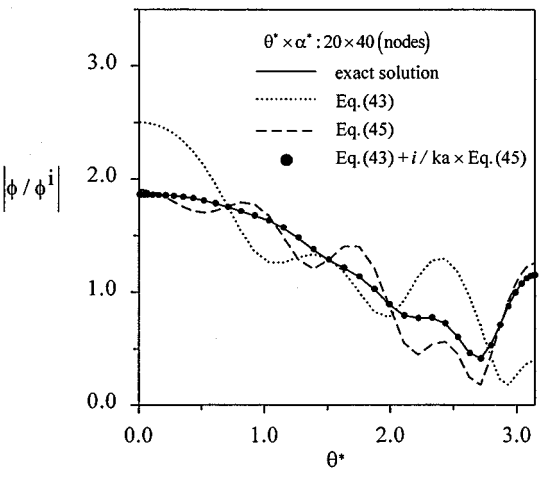


FIG. 4. Surface field for a hard sphere exposed to a plane wave with $ka = 4.4934$.

From Fig. 5 we observe that both numerical methods give good results at $ka = \pi$, 5π, and 10π. The present method still produces satisfactory results for the larger values of ka. However we notice that the Gauss–Legendre method does not converge in Fig. 5(d) and (e) for the higher wave numbers. The Kirchhoff approximation is also plotted in the figures for the comparison. The theory renders good estimations at higher wave numbers on the illuminated and shadowed sides. The discrepancy in the penumbra region is still quite pronounced at $ka = 20\pi$.

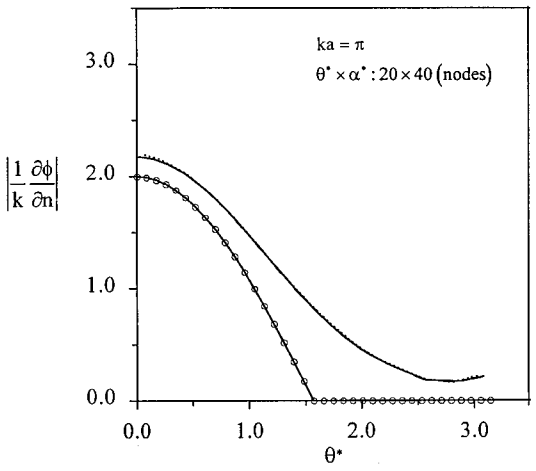
The scattering by a hard sphere is displayed in Fig. 6. In this case, the Gauss–Legendre method gives good result at the lowest wave number $ka = \pi$ only. For the other larger values of ka, the predictions by the Gauss–Legendre method are distinctly different from the exact ones. On the contrary, the present method can offer fairly reliable results up to $ka = 20\pi$. Figure 7 shows a steady improvement of the results by using the present method as the number of quadrature nodes increases. The deviations of the Gauss–Legendre method are still quite large. Figure 6 also reveals that the Kirchhoff approximation still can not estimate the scattering field satisfactorily at $ka = 20\pi$. This fact indicates that the application of Kirchhoff approximation to the hard bodies is more restrictive than that to the soft bodies.

At the end of this section we would like to mention that all of the selected values of ka are characteristic frequencies. The results presented above therefore have justified that the nonuniqueness and singularity troubles can be overcome successfully by using the formulations presented in the Secs. II and III.

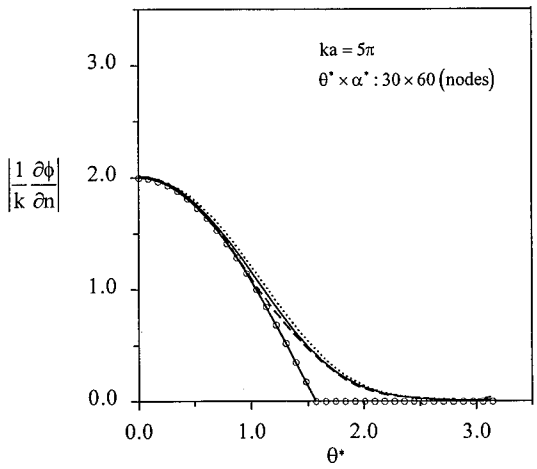
VI. CONCLUSIONS

A set of completely singularity-free integral formulations of the Helmholtz integral equation and its normal derivative has been constructed. The integrals involved in the resulting formulations can be integrated directly by using the standard quadrature methods. The validity of the formulations has been checked by solving the acoustic scattering from a soft or hard sphere, respectively. By extending the Filon's quadrature method, in conjunction with the singularity-free integral equations, for treating the highly oscillating integrands, fairly reliable results have been achieved up to wave number $ka = 20\pi$. The conventional Gauss–Legendre method yields satisfactory results at the lower values of ka only, as to be expected.

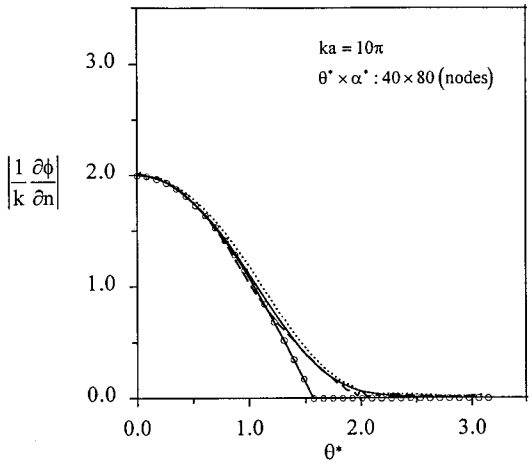
The present method has shown the capability of treating the acoustic scattering for a plane wave with a relatively wide wave number range $ka = \pi - 20\pi$. The discrepancy of the present numerical results from the exact solutions at the higher values of wave number may be due to the assumption of a constant distribution of unknown function along each integration subinterval. The remedy against the loss of accuracy may recourse to the adoption of higher-order approximations. It is hoped that the higher-order approximations could not only improve the accuracy, but also resolve the corresponding problems at even higher wave numbers.



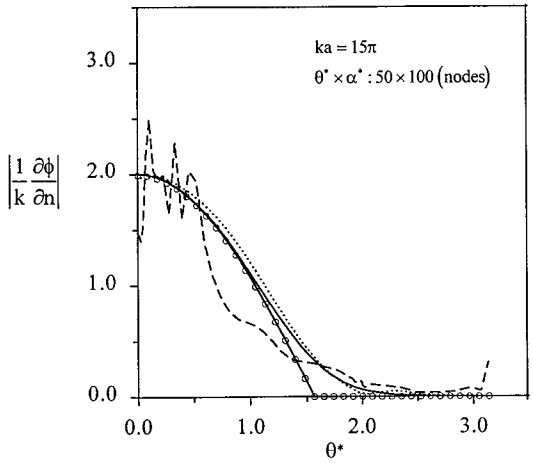
(a)



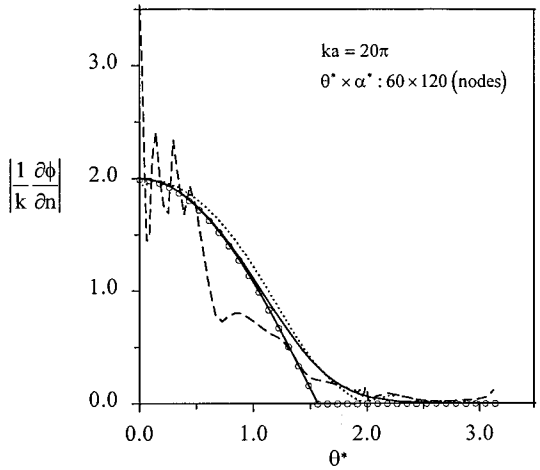
(b)



(c)

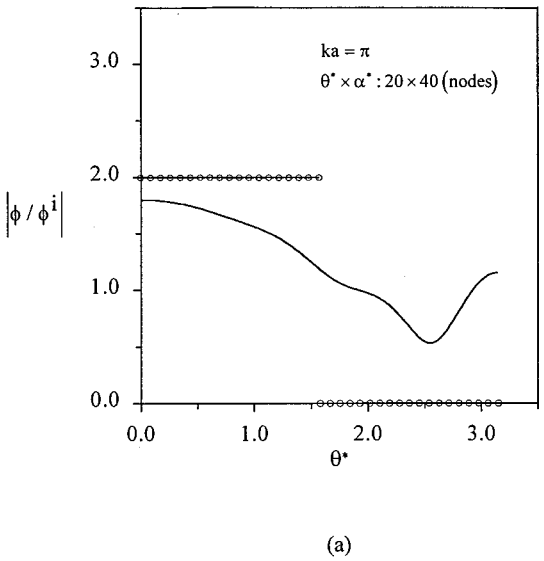


(d)

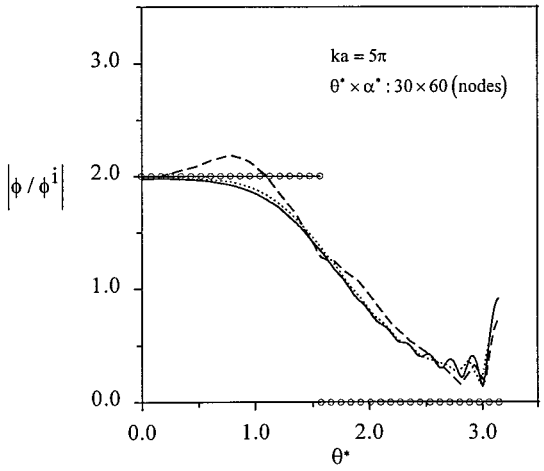


(e)

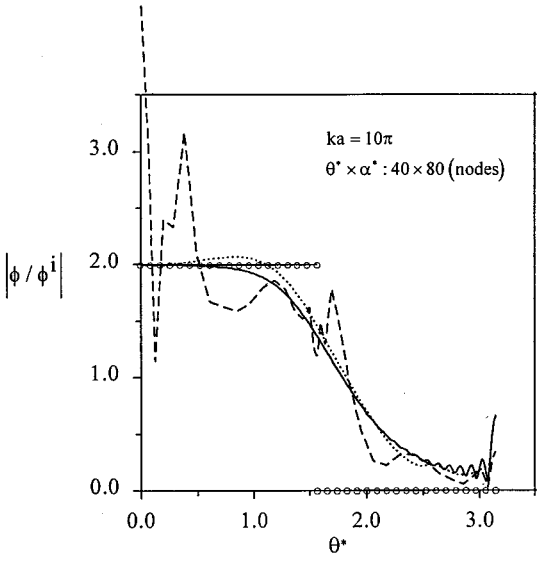
FIG. 5. Surface field for a soft sphere exposed to a plane wave with (a) $ka=\pi$, (b) $ka=5\pi$, (c) $ka=10\pi$, (d) $ka=15\pi$, (e) $ka=20\pi$ (solid line: analytical results; dotted line: present numerical results; dashed line: Gauss–Legendre method; $\ominus$: Kirchhoff approximation).



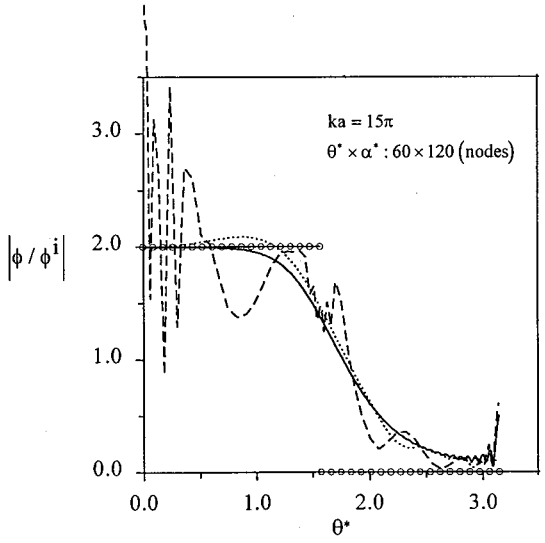
(a)



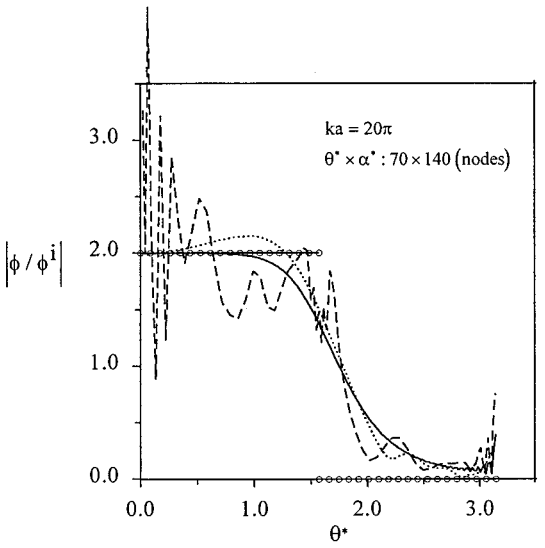
(b)



(c)

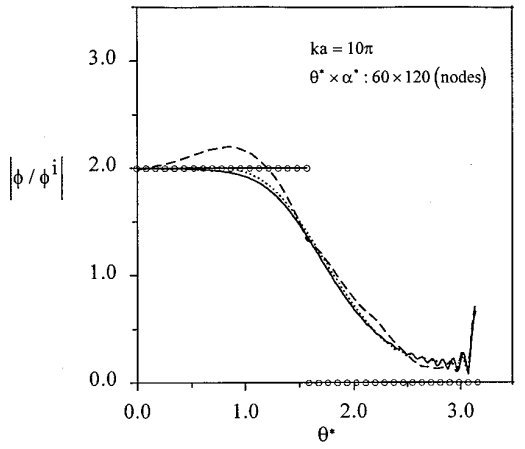


(d)

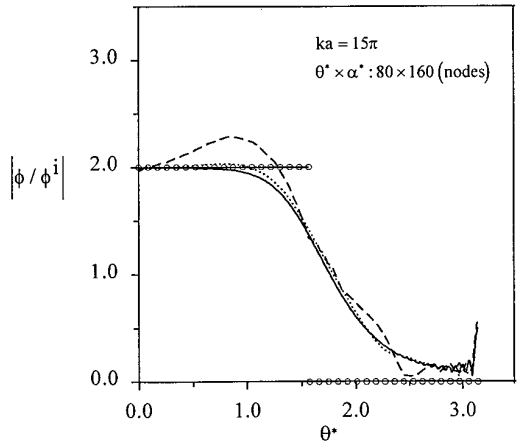


(e)

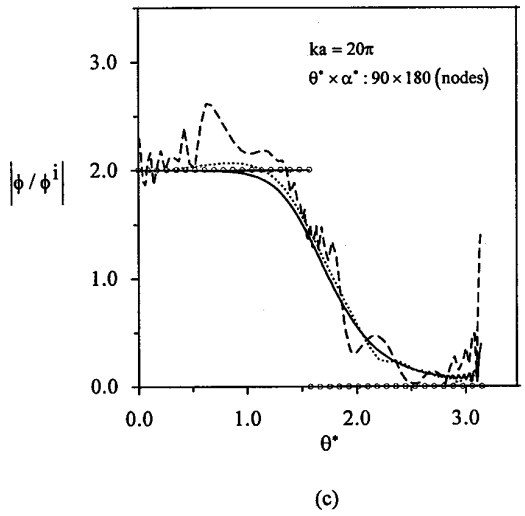
FIG. 6. Surface field for a hard sphere exposed to a plane wave with (a) $ka=\pi$, (b) $ka=5\pi$, (c) $ka=10\pi$, (d) $ka=15\pi$, (e) $ka=20\pi$ (solid line: analytical results; dotted line: present numerical results; dashed line: Gauss–Legendre method; ⊖: Kirchhoff approximation).



(a)



(b)



(c)

FIG. 7. Same as Fig. 6. with more integration nodes.

ACKNOWLEDGMENT

This work was partially supported by the National Science Council, R.O.C., under Grant No. NSC 86-2611-E-006-004.

[1] H. A. Schenck, ''Improved integral formulation for acoustic radiation problems,'' J. Acoust. Soc. Am. **44**, 41–58 (1968).

[2] D. T. Wilton, ''Acoustic radiation and scattering from elastic structures,'' Int. J. Numer. Methods Eng. **13**, 128–138 (1978).

[3] W. Tobocman, ''Calculations of acoustic wave scattering by means of the Helmholtz integral equation I,'' J. Acoust. Soc. Am. **76**, 599–607 (1984).

[4] W. Tobocman, ''Extension of the Helmholtz integral equation method to shorter wavelengths,'' J. Acoust. Soc. Am. **80**, 1828–1837 (1986).

[5] A. F. Seybert, B. Soenarko, F. J. Rizzo, and D. J. Shippy, ''An advanced computational method for radiation and scattering for acoustic waves in three dimensions,'' J. Acoust. Soc. Am. **77**, 362–368 (1985).

[6] A. F. Seybert, B. Soenarko, F. J. Rizzo, and D. J. Shippy, ''A special integral equation formulation for acoustic radiation and scattering for axisymmetric bodies and boundary conditions,'' J. Acoust. Soc. Am. **80**, 1241–1247 (1986).

[7] A. F. Seybert and T. K. Rengarajan, ''The use of CHIEF to obtain unique solutions for acoustic radiation using boundary integral equations,'' J. Acoust. Soc. Am. **80**, 1299–1306 (1987).

[8] A. J. Burton and G. F. Miller, ''The application of integral equation methods to the numerical solution of some exterior boundary value problems,'' Proc. R. Soc. London, Ser. A **323**, 201–210 (1971).

[9] W. S. Hall and W. H. Robertson, ''Standard Helmholtz integral equation calculations near characteristic frequencies,'' J. Sound Vib. **126**, 367–368 (1988).

[10] A. F. Seybert, ''A note on methods for circumventing nonuniqueness when using integral equations,'' J. Sound Vib. **115**, 171–172 (1987).

[11] Z. Reut, ''On the boundary integral methods for the exterior acoustic problem,'' J. Sound Vib. **103**, 297–298 (1985).

[12] David T. I. Francis, ''A gradient formulation of the Helmholtz integral equation for acoustic radiation and scattering,'' J. Acoust. Soc. Am. **93**, 1700–1709 (1993).

[13] W. L. Meyer, W. A. Bell, B. T. Zinn, and M. P. Stallybrass, ''Boundary integral solutions of three dimensional acoustic radiation problems,'' J. Sound Vib. **59**, 245–262 (1978).

[14] I. C. Mathews, ''Numerical techniques for three-dimensional steady-state fluid-structure interaction,'' J. Acoust. Soc. Am. **79**, 1317–1325 (1986).

[15] G. Krishnasamy, L. W. Schmerr, T. J. Rudolphi, and F. J. Rizzo, ''Hypersingular boundary integral equations: Some applications into acoustic and elastic wave scattering,'' J. Appl. Mech. **57**, 404–414 (1990).

[16] Yijun Liu and F. J. Rizzo, ''A weakly singular form of the hypersingular boundary integral equation applied to 3-D acoustic wave problems,'' Comput. Methods Appl. Mech. Eng. **96**, 271–287 (1992).

[17] C. C. Chien, H. Rajiyah, and S. N. Atluri, ''An effective method for solving the hypersingular integral equations in 3-D acoustics,'' J. Acoust. Soc. Am. **88**, 918–937 (1990).

[18] W. S. Hwang, ''Hypersingular boundary integral equation for acoustic problems,'' J. Acoust. Soc. Am. **101**, 3336–3342 (1997).

[19] R. Leis, ''Zur Dirichletschen Randwertaufgabe des Außenraumes der Schwingungsgleichung,'' Math. Z. **90**, 205–211 (1965).

[20] O. I. Panich, ''On the question of the solvability of the exterior boundary problem for the wave equation and Maxwell's equation (in Russian),'' Usp. Mat. Nauk. **20**, 221–226 (1965).

[21] H. Brakhage and P. Werner, ''Über das Dirichletsche Aussenraumproblem für die helmholtzsche Schwingungsgleichung,'' Arch. Math. **16**, 325–329 (1965).

[22] A. J. Burton, ''The solution of Helmholtz equation in exterior domains using integral equations,'' NPL Report NAC 30, National Physical Laboratory, January (1973).

[23] W. Tobocman, ''Extension of the Helmholtz integral equation method to shorter wavelengths II,'' J. Acoust. Soc. Am. **82**, 704–706 (1987).

[24] L. N. G. Filon, ''On a quadrature formula for trigonometric integrals,'' Proc. R. Soc. Edinburgh **49**, 38–47 (1928).

[25] Y. L. Luke, ''On the computation of oscillatory integrals,'' Proc. Camb. Philos. Soc. **50**, 269–277 (1954).

[26] A. M. O. Smith, ''Numerical integration of oscillating functions having a nonlinear argument,'' Aircraft Division, Douglas Aircraft Company, Long Beach, California, Engineering Paper 3373 (1967).

[27] H.-T. Shen and C. Farell, ''Numerical calculation of the wave integrals in the linearized theory of water waves,'' J. Ship Res. **21**, 1–10 (1977).

[28] J. C. Lachat and J. O. Watson, ''Effective numerical treatment of boundary integral equations: A formulation for three-dimensional elastostatics,'' Int. J. Numer. Methods Eng. **10**, 991–1005 (1976).

[29] L. Landweber and M. Macagno, ''Irrotational flow about ship forms,'' IIHR Report No. 123, Iowa Institute of Hydraulic Research, The University of Iowa (1969).

Nonlinear acoustic waves in porous media in the context of Biot's theory

D. M. Donskoy, K. Khashanah, and T. G. McKee, Jr.
Stevens Institute of Technology, Hoboken, New Jersey 07030

(Received 22 July 1996; revised 26 February 1997; accepted 2 June 1997)

The nonlinear dynamic equations introduced by Biot to model porous media have not been implemented to describe nonlinear acoustic waves in such media. In this work the equations are revised and a mathematical model depicting the physical nonlinearity is established. A perturbation technique is then applied to find solutions to the nonlinear Biot equations. An important feature of the developed model is the introduction of the dependence of the structural parameters of the medium on its porosity. The model establishes a correlation between the measurable effective nonlinear parameter and structural parameters of the porous medium. This suggests employing nonlinear measurements as a diagnostic tool for porous media. © *1997 Acoustical Society of America.* [S0001-4966(97)01211-3]

PACS numbers: 43.25.−x [MAB]

INTRODUCTION

Nonlinear dynamic behavior of porous media, natural or artificial, has attracted increasing attention recently. It has been observed experimentally that porous media, such as soil, rocks, sediments, etc., exhibit a strong elastic nonlinearity, in some cases two to three orders of magnitude greater as compared with nonporous media.[1–8] Although a considerable number of experimental studies of nonlinear acoustic phenomena in porous media have been conducted, the theoretical investigation of these phenomena is in its beginnings.

Common approaches to describe the nonlinear dynamic effects in porous media include: the five-constant elastic theory,[9] developed for isotropic nonporous solids which approximates a porous medium by an elastic nonporous medium with effective linear and nonlinear parameters;[6,8] the bimodular and hysteresis models;[8,10] and the homogenization technique.[11]

However, these approaches do not clearly exhibit the relationship between the measurable nonlinear effects and real physical parameters of the medium such as porosity, elastic moduli of a frame, solid, and liquid/gaseous phases of the porous medium. In addition, these methods do not take into account the dynamic behavior of the liquid/gas filling the pores which reduces the number of longitudinal types of waves (fast and slow waves) to just one.

The studies[12,13] developed relationships between the nonlinear parameter and the porosity based on the nonlinear dynamic behavior of a single spherical or cylindrical pore. However, the results are limited to media with isolated spherical/cylindrical pores and very low porosity (10^{-4} and lower).

The most realistic model of poroelastic media was developed by Biot.[14] Biot's theory is widely used in mechanics as well as in linear acoustics. It has proved to be the most accurate description of the static and dynamic behavior of porous media. Biot proposed the theory of finite (nonlinear) deformations of porous solids,[15] in which the third-order stress–strain relationships for the liquid-saturated porous media are derived.

This work extends the Biot theory in two directions: first, by deriving nonlinear equations for acoustic wave propagation in porous media and second, by establishing relationships between measurable nonlinear parameters and the physical characteristics of the porous media.

The structure of this paper is as follows: in the first section, the extension of Biot's theory is made; in the second section, these extensions are applied to the one-dimensional case; and in the third section, the nonlinear wave interactions are calculated. Section IV justifies incorporating a porosity dependence in the structural parameters of the medium and introduces the effective nonlinear parameter.

I. THE SEMILINEAR BIOT MODEL

As Biot pointed out,[15] ''strong nonlinear behavior is associated with the microstructure of the pores. For example, partial pore closure on changes in contact areas will generally cause appreciable nonlinear dependence of effective stress on average strain of the bulk material, even for strains of the order 1 or 2%.'' At the same time, ''volume changes of the grains in the solid matrix and fluid in the pore are approximately linear dependent on stresses and pressures.'' This physical mechanism of nonlinear behavior of porous media (assuming linear behavior of the material of solid matrix and pore fluid) fundamentally brings out the semilinear model.

In terms of the number of coefficients needed to characterize a nonlinear medium, the semilinear model requires a total of seven constants, four to characterize the linear behavior and three to characterize the nonlinear behavior. As implied by Eqs. (26) and (29) on page 167 of Ref. 14, a total of 11 constants are needed if semilinearity is not assumed. For such a model, see Ref. 16.

Using the semilinear model, Biot derived nonlinear stress–strain relations coupled with Darcy's law. However, the equations of motion that describe nonlinear acoustic wave propagation were not developed. These equations will be derived later in this section.

First some notation is needed.

A. General notation

Throughout this paper we use boldface capital letters for matrices, boldface lower case for vectors, and lower case for scalars. ∂_x means partial differentiation with respect to x. The tensor summation notation applies for repeated indices. Prime is only used to denote differentiation.

B. Functions and variables

$\mathbf{u}$	the solid displacement vector
$\mathbf{w}$	the fluid displacement vector relative to the solid
ϵ_{ij}	the solid strain
ζ	$=-\text{div}\ \mathbf{w}$ the volumetric fluid content
σ_{ij}	the stress tensor
p	the fluid pressure
H	the elastic energy invariant
x_i	the ith component of the position vector $\mathbf{x}=(x_1,x_2,x_3)$ in space
t	time

C. Constants and coefficients

β	porosity
ν	Poisson ratio
μ	shear modulus of frame
K_s	bulk modulus of frame
K_f	bulk modulus of fluid
K_r	bulk modulus of grains
μ_r	shear modulus of grains
ρ	density of grains
ρ_f	density of fluid
η	fluid viscosity
κ	permeability
D,F,G	nonlinear coefficients
Γ	effective nonlinear parameter
D_r	$=K_r[1+\beta(K_r/K_f-1)]$
λ	$=K_s-\frac{2}{3}\mu$ Lamé coefficient of frame
α	$=1-K_s/K_r$ Biot's elastic coefficient
M	$=K_r^2/(D_r-K_s)$ Biot's elastic coefficient
m	$=(\rho_f/\beta)(1+a)$ Darcy's coefficient
a	added mass coefficient
θ	$=1/3K_r$ fluid compliance coefficient (Biot uses β)

We now derive the equations of motion for the displacement in an isotropic semilinear porous medium by first establishing the strain–displacement relations, the stress–strain relations, and then substituting them in the equations of motion.

The semilinear model neglects the geometric nonlinear terms $\partial_{x_i}u_k\cdot\partial_{x_j}u_k$ in the strain–displacement relationships:

$$\epsilon_{ij}=\tfrac{1}{2}(\partial_{x_j}u_i+\partial_{x_i}u_j). \tag{1}$$

To obtain the stress–strain relations we define the strain invariants

$$I_1=\epsilon_{11}+\epsilon_{22}+\epsilon_{33}, \tag{2}$$

$$I_2=\epsilon_{22}\epsilon_{33}+\epsilon_{33}\epsilon_{11}+\epsilon_{11}\epsilon_{22}-\epsilon_{23}\epsilon_{32}-\epsilon_{31}\epsilon_{13}-\epsilon_{12}\epsilon_{21}, \tag{3}$$

$$I_3=\det(\epsilon_{ij}). \tag{4}$$

Following Biot,[15] we introduce a more convenient set of invariants in terms of the ones given above [note that there is an error in Eq. (28)[15]]

$$U=\tfrac{1}{3}I_1^3-I_1I_2+I_3, \tag{5}$$

$$B=I_1I_2-3I_3. \tag{6}$$

When the effect of the fluid pressure is considered, one obtains, as in Ref. 15,

$$\overline{\epsilon_{ij}}=\epsilon_{ij}+\theta\delta_{ij}p, \tag{7}$$

where θ is a compliance constant (note that Biot uses β in place of θ) and p is the fluid pressure increase. Let us explain briefly the significance of θ according to Biot.[15] Let v be the volume increase in the pores per unit element and let $\epsilon=\delta_{ij}\epsilon_{ij}$ be the increase in the bulk volume. Then the volume change v_s of the solid matrix is $v_s=\epsilon-v$. The assumption of semilinearity implies a linear relation between v_s and the stress $\overline{\sigma_{ij}}$. Hence the linear relation

$$v_s=\theta_{ij}\overline{\sigma_{ij}}$$

is used with, generally, six compliance constants. If the medium is isotropic, as is assumed here, $\theta_{ij}=\delta_{ij}\theta$. The strain function can now be defined as

$$H=D\overline{U}+F\overline{B}+G\overline{I_3}, \tag{8}$$

where $\overline{U}$, $\overline{B}$, and $\overline{I_3}$ are obtained from the definitions above by replacing the terms ϵ_{ij} with corresponding $\overline{\epsilon_{ij}}$. The constants D, F, and G are coefficients that characterize the nonlinear behavior of an isotropic porous medium. The rate of change of the function H with respect to $\overline{\epsilon_{ij}}$ accounts for the nonlinearity in the expression of the stress σ_{ij} in terms of the strain $\overline{\epsilon_{ij}}$. In view of the fact that we have found a missing coefficient in Biot's Eqs. (76),[15] we reproduce the derivation of these equations. Keep in mind that θ is used here in place of β in Biot's work.

We start from Eqs. (67),[15]

$$\begin{aligned}\sigma_{ij}&=\overline{\sigma_{ij}}(\overline{\epsilon_{lk}})-\delta_{ij}p,\\ \zeta&=\epsilon-\theta_{ij}\overline{\sigma_{ij}}(\overline{\epsilon_{lk}})+C_ap,\end{aligned} \tag{9}$$

where C_a is a mixed fluid–solid compressibility. According to Eq. (71)[15] we have

$$\overline{\sigma_{ij}}=2\mu\overline{\epsilon_{ij}}+\delta_{ij}\lambda\overline{I_1}+\partial_{\overline{\epsilon_{ij}}}H, \tag{10}$$

where λ and μ are the Lamé coefficients and $\overline{\sigma_{ij}}$ refers to the stress in the absence of fluid pressure (the jacketed unit element under stresses $\overline{\sigma_{ij}}$ applied to the faces of the unit element while the pore fluid pressure is maintained at constant atmospheric pressure).

Substitute (7) into (10) and take into consideration that $\theta_{ij}=\delta_{ij}\theta$ and we get the equation

$$\begin{aligned}\sigma_{ij}&=2\mu\epsilon_{ij}+\delta_{ij}\lambda\epsilon+\delta_{ij}[\theta(2\mu+3\lambda)-1]p+\partial_{\overline{\epsilon_{ij}}}H,\\ \zeta&=[1-\theta(2\mu+3\mu)]\epsilon+[C_a-3\theta^2(2\mu+3\lambda)]p\\ &\quad-\delta_{ij}\theta\,\partial_{\overline{\epsilon_{ij}}}H.\end{aligned} \tag{11}$$

Therefore with

$$\frac{1}{M}=C_a-3\theta^2(2\mu+3\lambda),$$

$$\alpha=1-\theta(2\mu+3\lambda)$$

the following equations are derived:

$$\begin{aligned}\sigma_{ij}&=2\mu\epsilon_{ij}+\delta_{ij}(\lambda\epsilon-\alpha p)+\partial_{\overline{\epsilon}_{ij}}H,\\ \zeta&=\alpha\epsilon+\frac{1}{M}\,p-\delta_{ij}\theta\;\partial_{\overline{\epsilon}_{ij}}H.\end{aligned}\tag{12}$$

The factor θ, the last term of the second equation, is missing in Biot's Eqs. (76).[15] Notice that dimensionally, θ has reciprocal-of-pressure units and $\partial_{\overline{\epsilon}_{ij}}H$ pressure units. Thus as is ζ, the term $\theta\partial_{\overline{\epsilon}_{ij}}H$ is dimensionless.

From the second equation, we get the pressure p

$$p=M(\zeta-\alpha\epsilon)+\delta_{ij}\theta M\;\partial_{\overline{\epsilon}_{ij}}H.\tag{13}$$

In these equations α and M are Biot coefficients of the linear theory of porous media. We mention briefly here that the constant α measures the ratio of the fluid volume forced out to the volume change of the porous medium if the medium is compressed while allowing the fluid to escape (the stress is zero). The constant M is the reciprocal of the fluid amount which can be forced into the porous medium under pressure while the volume of the medium is kept constant.[14] In the limiting case where the fluid is incompressible and the elastic matrix of the porous medium is also incompressible, α and M become $\alpha=1$ and $M=\infty$. It is shown by Biot that α is such that $\beta\leqslant\alpha\leqslant1$, where β is the porosity of the medium. A more precise estimate is provided in Ref. 17.

Next, we will eliminate the explicit fluid pressure dependence from the stress–strain relationship. Let us write the stress tensor and the fluid pressure as

$$\sigma_{ij}=\sigma_{ij}^l+\sigma_{ij}^n,\quad p=p^l+p^n,\tag{14}$$

where the functions σ_{ij}^l and p^l are linear functions and σ_{ij}^n and p^n are nonlinear functions of the strain and ζ. More explicitly,

$$\begin{aligned}&\sigma_{ij}^l=2\mu\epsilon_{ij}+\delta_{ij}(\lambda\epsilon-\alpha p),\quad \sigma_{ij}^n=\partial_{\overline{\epsilon}_{ij}}H,\\ &p^l=M(\zeta-\alpha\epsilon),\quad p^n=\delta_{ij}\theta M\;\partial_{\overline{\epsilon}_{ij}}H.\end{aligned}\tag{15}$$

The nonlinear part of the stress tensor is obtained, once the term $\partial_{\overline{\epsilon}_{ij}}H$ is calculated. It is clear that $\partial_{\overline{\epsilon}_{ij}}I_1=\delta_{ij}$ while $\partial_{\overline{\epsilon}_{ij}}\overline{I_2}=\delta_{ij}I_1-\overline{\epsilon_{ij}}$ and $\partial_{\overline{\epsilon}_{ij}}\overline{I_3}=\mathrm{cof}(\overline{\epsilon_{ij}})$, where cof denotes the cofactor of the element $\overline{\epsilon_{ij}}$. Thus the nonlinear expression becomes

$$\begin{aligned}\partial_{\overline{\epsilon}_{ij}}H=&\,\delta_{ij}\overline{I_1^2}-(F-D)(\overline{I_1\epsilon_{ij}}-\delta_{ij}\overline{I_2})\\ &+(D-3F+G)\mathrm{cof}(\overline{\epsilon_{ij}}).\end{aligned}$$

Each of these terms is a second-order term involving the strain tensor $\overline{\epsilon_{ij}}$. This presents a problem since, in view of relation (7), it is seen that the second equation in (12) is a quadratic equation in p that implicitly defines the fluid pressure in terms of the strain. Therefore it is not possible to eliminate p from the first equation in (12). In keeping with the second-order approximation, the problem is circumvented by substituting the linear part of the pressure, $p=p^l$, into the nonlinear expression of the stress, σ_{ij}^n, and substituting the linear and nonlinear part of the pressure, $p=p^l+p^n$, into the linear part of the stress, σ_{ij}^l. Therefore the constitutive relations (12) become

$$\begin{aligned}\sigma_{ij}=&\,2\mu\epsilon_{ij}+\delta_{ij}[(\lambda+\alpha^2M)\epsilon-\alpha M\zeta]+\partial_{\epsilon_{ij}}\overline{H}\big|_{p=p^l}\\ &-\alpha\theta M\delta_{ij}\delta_{lm}\partial_{\overline{\epsilon}_{lm}}H\big|_{p=p^l},\\ p=&\,M(\zeta-\alpha\epsilon)+\delta_{ij}\theta M\partial_{\overline{\epsilon}_{ij}}H\big|_{p=p^l}.\end{aligned}\tag{16}$$

Let the vectors $\mathbf{u}:=(u_1,u_2,u_3)^T$ and $\mathbf{w}:=(w_1,w_2,w_3)^T$ denote the solid displacement and the fluid displacement relative to the solid, respectively. Then the general form of the equations of motion, according to Biot's theory,[14] are

$$\partial_{x_j}\sigma_{ij}=\rho\;\partial_{tt}u_i+\rho_f\;\partial_{tt}w_i,\tag{17}$$

$$-\partial_{x_i}p=\rho f\;\partial_{tt}u_i+\;Y\;\partial_t w_i.\tag{18}$$

Here $Y=\eta/k+m\partial_t$ is a viscodynamic operator. If the second term in Y is suppressed (zero frequency), Eq. (18) reduces to Darcy's law.

It is clear that if, in Eqs. (12), the nonlinear terms are suppressed and the resulting linear constitutive equations are substituted in the equations of motion, the familiar linear Biot equations involving the displacements are obtained.

The equations of motion involving the displacement vectors in their general form are now obtained by substituting the strain–displacement relations (1) and (7) into the constitutive equations (16) and finally into the equations of motion [(17) and (18)]. In order to obtain a more presentable result, we will explicitly write out these equations for the one-dimensional case and devise a method for solving those equations. The analysis carried out in the next section can be extended to equations in higher dimensions.

II. THE ONE-DIMENSIONAL CASE

Let u and w be the solid displacement and the fluid displacement relative to the solid, respectively. Both displacements are along the x axis. In this case the equations of motion (17) and (18) reduce to two coupled second-order nonlinear partial differential equations with constant coefficients. In what follows it is shown that the linear part of the coupled system corresponds to the linear Biot equations of motion while the nonlinear part appears as second-order terms involving the spatial derivatives of the displacements as desired. It is also shown, as expected and known from the linear Biot theory, that two compressional waves exist as solutions to the linear equations.

The equations of motion are as follows

$$\begin{aligned}&(\lambda_c+\alpha^2M)u_{xx}+\alpha Mw_{xx}\\ &\quad=\rho u_{tt}+\rho_f w_{tt}+N_1(u_x,u_{xx},w_x,w_{xx}),\\ &\alpha Mu_{xx}+Mw_{xx}=\rho_f u_{tt}+Yw_t+N_2(u_x,u_{xx},w_x,w_{xx}),\end{aligned}\tag{19}$$

where $\lambda_c=\lambda+2\mu$ and N_1 and N_2 are nonlinear functions which will be defined subsequently. It is immediate from examining Eqs. (19) that the linear part corresponds to the

linear Biot equations of motion. The viscodynamic operator Y can be considered in two separate cases; in the first case the nondissipative system, in which $Y:=m\partial_t$, is investigated here. In the second case, the dissipative system $Y:=m\partial_t+\eta/\kappa$ will be studied in a later paper. The nonlinear functions N_1 and N_2 involve the nonlinear coefficients of the medium according to

$$\begin{aligned} N_1&=c_{11}u_xu_{xx}+c_{12}u_xw_{xx}+d_{11}w_xu_{xx}+d_{12}w_xw_{xx},\\ N_2&=c_{21}u_xu_{xx}+c_{22}u_xw_{xx}+d_{21}w_xu_{xx}+d_{22}w_xw_{xx}. \end{aligned} \tag{20}$$

Let $K:=D+6F+G$ and $P:=D+2F$, then

$$\begin{aligned} c_{11}&=2(\alpha-1)(\alpha^2\beta^2M^2K-2\alpha\beta MP+D),\\ c_{12}&=2\beta M(\alpha-1)(\alpha\beta MK-P),\\ c_{21}&=2M(3\alpha^2\beta^2M^2K-2\alpha\beta MK+P),\\ c_{22}&=2\beta M^2K(3\alpha\beta M-1),\\ d_{11}&=c_{12},\quad d_{12}=2\beta^2M^2K(\alpha M-1),\\ d_{21}&=c_{22},\quad d_{22}=6\beta^2M^3K. \end{aligned} \tag{21}$$

In order to further facilitate our calculations, we define the matrices $\mathbf{C}:=(c_{ij})$ and $\mathbf{D}:=(d_{ij})$ and the displacement vector $\mathbf{v}:=(u,w)^T$. Then the nonlinear operations expressed by (20) can be rewritten in terms of the operator N defined by

$$N[\mathbf{v}]=[\mathbf{C}u_x+\mathbf{D}w_x]\mathbf{v}_{xx}. \tag{22}$$

Define the linear nondissipative wave operator L as

$$L[\mathbf{v}]=(\mathbf{A}\partial_{xx}-\mathbf{R}\partial_{tt})\mathbf{v}, \tag{23}$$

where in this case the matrices $\mathbf{A}:=(a_{ij})$ and $\mathbf{R}:=(r_{ij})$ are as follows: $a_{11}=\lambda_c+\alpha^2M$, $a_{12}=a_{21}=\alpha M$, $a_{22}=M$, $r_{11}=\rho$, $r_{12}=r_{21}=\rho_f$, and $r_{22}=m$. Equations (19) become

$$L[\mathbf{v}]=N[\mathbf{v}]. \tag{24}$$

III. WAVE SOLUTIONS

The first step in the solution is to find wave solutions to the linear equation

$$L[\mathbf{v}]=0. \tag{25}$$

Let k_1 and k_2 be the roots of the characteristic equation

$$\det[\omega^2\mathbf{R}-k^2\mathbf{A}]=0 \tag{26}$$

associated with the operator L. Written explicitly, the characteristic equation is

$$\begin{aligned} &\lambda_cMk^4+[(2\alpha\rho_f-\rho)M-m(\lambda_c+\alpha^2M)]\omega^2k^2\\ &\quad+(\rho m-\rho_f^2)\omega^4=0. \end{aligned} \tag{27}$$

Note that this characteristic equation defines a nondispersive relation between ω and k. The linear wave solution is

$$\mathbf{v}_L(x,t)=\mathbf{a}_1v_1+\mathbf{a}_2v_2, \tag{28}$$

where the amplitudes $\mathbf{a}_1=(a_1,a_3)^T$ and $\mathbf{a}_2=(a_2,a_4)^T$ are constant and $v_1=e^{i(\omega t-k_1x)}$ and $v_2=e^{i(\omega t-k_2x)}$. To find the second-harmonic waves from the linear wave solutions, we seek wave solutions with slowly varying amplitudes that have the form

$$\mathbf{v}_N(x,t)=\mathbf{b}_1(x)v_1^2+\mathbf{b}_2(x)v_2^2+\mathbf{b}_3(x)v_1v_2, \tag{29}$$

in which the functions $\mathbf{b}_1(x)=(b_1(x),b_4(x))^T$, $\mathbf{b}_2=(b_2(x),b_5(x))^T$, and $\mathbf{b}_3=(b_3(x),b_6(x))^T$ are to be determined. The linear wave solution now is operated upon by the nonlinear operator N and the nonlinear waves are acted upon by the linear operator L:

$$L[\mathbf{v}_N]=N[\mathbf{v}_L]. \tag{30}$$

Matching coefficients of v_1^2, v_2^2, and v_1v_2 and using the linear independence of the wave solutions, the following second-order systems of differential equations are obtained:

(1) From the coefficient of v_1^2

$$\begin{aligned} &\mathbf{A}[\mathbf{b}_1''(x)-4ik_1\mathbf{b}_1'(x)]+4(\omega^2\mathbf{R}-k_1^2\mathbf{A})\mathbf{b}_1\\ &\quad=ik_1^3(a_1\mathbf{C}+a_3\mathbf{D})\mathbf{a}_1. \end{aligned}$$

(2) From the coefficient of v_2^2

$$\begin{aligned} &\mathbf{A}[\mathbf{b}_2''(x)-4ik_2\mathbf{b}_2'(x)]+4(\omega^2\mathbf{R}-k_2^2\mathbf{A})\mathbf{b}_2(x)\\ &\quad=ik_2^3(a_2\mathbf{C}+a_4\mathbf{D})\mathbf{a}_2. \end{aligned}$$

(3) From the coefficient of v_1v_2

$$\begin{aligned} &\mathbf{A}[\mathbf{b}_3''(x)-2i(k_1+k_2)\mathbf{b}_3'(x)]+[4\omega^2\mathbf{R}-(k_1+k_2)^2\mathbf{A}]\mathbf{b}_3(x)\\ &\quad=i[k_2k_1^2(a_2\mathbf{C}+a_4\mathbf{D})\mathbf{a}_1+k_1k_2^2(a_1\mathbf{c}+a_3\mathbf{D})\mathbf{a}_2]. \end{aligned}$$

In view of the characteristic equation, the first two systems are degenerate in the sense that the matrix coefficients of $\mathbf{b}_j$, $j=1,2$ in both systems are singular. This is also an indication that an attempt to find second harmonic wave solutions with constant coefficients, using this technique, must fail. The third system, however, is nondegenerate.

Solutions to these systems are obtained using standard procedures. However, for our purpose we consider slowly varying second-harmonic amplitudes in the sense that the second derivatives of $\mathbf{b}_j$, $j=1,2,3$ are negligible. With this approximation, the following systems of first-order equations are to be solved

(1) For the coefficient of v_1^2

$$\begin{aligned} \mathbf{b}_1'(x)=&\frac{\mathbf{A}^{-1}}{ik_1}(\omega^2\mathbf{R}-k_1^2\mathbf{A})\mathbf{b}_1(x)-\frac{k_1^2\mathbf{A}^{-1}}{4}\\ &\times(a_1\mathbf{C}+a_3\mathbf{D})\mathbf{a}_1. \end{aligned}$$

(2) For the coefficient of v_2^2

$$\begin{aligned} \mathbf{b}_2'(x)=&\frac{\mathbf{A}^{-1}}{ik_2}(\omega^2\mathbf{R}-k_2^2\mathbf{A})\mathbf{b}_2(x)-\frac{k_2^2\mathbf{A}^{-1}}{4}\\ &\times(a_2\mathbf{C}+a_4\mathbf{D})\mathbf{a}_2. \end{aligned}$$

(3) For the coefficient of v_1v_2

$$\begin{aligned} \mathbf{b}_3'(x)=&\frac{\mathbf{A}^{-1}}{2i(k_1+k_2)}\{[4\omega^2\mathbf{R}-(k_1+k_2)^2\mathbf{A}]\mathbf{b}_3(x)\\ &-i[k_2k_1^2(a_2\mathbf{C}+a_4\mathbf{D})\mathbf{a}_1+k_1k_2^2(a_1\mathbf{C}\\ &+a_3\mathbf{D})\mathbf{a}_2]\}. \end{aligned}$$

The matrix $\mathbf{A}$ is nonsingular since $\det\mathbf{A}=\lambda+2\mu\neq0$. The general form of these systems is

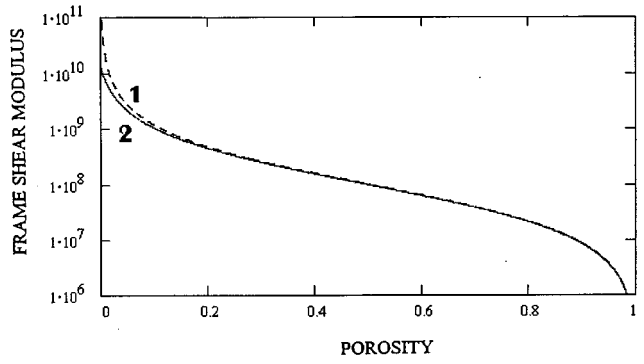


FIG. 1. Shear modulus of frame versus porosity. (1) corresponds to Eq. (33), (2) corresponds to the modified Eq. (34).

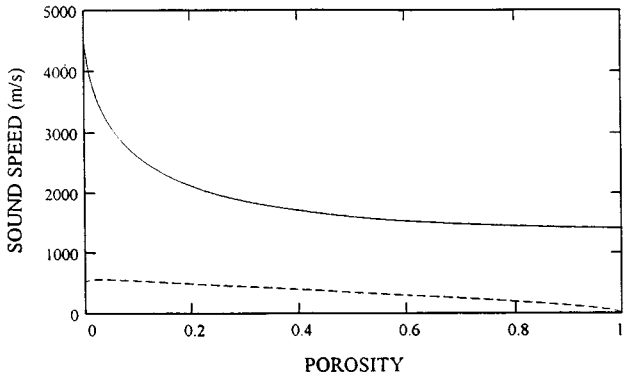


FIG. 2. Sound speeds versus porosity for a fluid-saturated porous medium: c_1-dashed line, c_2-solid line.

$$\mathbf{y}'(x)=\mathbf{B}\mathbf{y}(x)+\mathbf{f},$$

where $\mathbf{B}:=(b_{ij})$ is a constant coefficient matrix and $\mathbf{f}:=(f_1,f_2)^T$ is a constant forcing term.

In the first two systems the coefficient matrix is singular which means that one of the eigenvalues is zero and the other is the trace of the matrix **B**. The solution is composed of two parts, the homogeneous solution (unforced response) and the particular solution. The arbitrary constants in the homogeneous solution are fixed by the condition $\mathbf{y}(0)=0$. The solution is then given by

$$\mathbf{y}(x)=\frac{1}{\kappa}(b_{22}f_1-b_{12}f_2)x\xi^{(1)}+\frac{1}{\kappa^2}(b_{11}f_1+b_{12}f_2)\times(e^{\kappa x}-1)\xi^{(2)}, \tag{31}$$

where $\xi^{(1)}=(b_{12},-b_{11})^T$ and $\xi^{(2)}=(b_{12},b_{22})^T$ are eigenvectors corresponding to the eigenvalues 0 and $b_{11}+b_{22}$, respectively. $\kappa=b_{11}+b_{22}$ is used in the above formula. In the first two systems under consideration, the matrix **B** has the elements

$$\begin{aligned}
b_{11}&=-\frac{i}{k}\left[\frac{\omega^2}{\lambda_c}(\rho-\alpha\rho_f)-k^2\right],\\
b_{12}&=-\frac{i\omega^2}{k\lambda_c}(\rho_f-\alpha m),\\
b_{21}&=-\frac{i\omega^2}{k\lambda_c}\left[\frac{\rho_f}{M}(\lambda_c+\alpha^2M)-\rho\alpha\right],\\
b_{22}&=-\frac{i}{k}\left\{\frac{\omega^2}{\lambda_c}\left[\frac{m}{M}(\lambda_c+\alpha^2M)-\rho_f\alpha\right]-k^2\right\}.
\end{aligned} \tag{32}$$

In these expressions, use $k=k_1$ to calculate $\mathbf{b}_1$ and $k=k_2$ to calculate $\mathbf{b}_2$. Let us write out the solution for $\mathbf{b}_1$, the coefficient of the nonlinear wave v_1^2. The solution is

$$b_1(x)=\frac{1}{\kappa}(b_{22}f_1-b_{12}f_4)xb_{12}+\frac{1}{\kappa^2}(b_{11}f_1+b_{12}f_4)\times(e^{\kappa x}-1)b_{12},$$

$$b_4(x)=-\frac{1}{\kappa}(b_{22}f_1-b_{12}f_4)xb_{11}+\frac{1}{\kappa^2}(b_{11}f_1+b_{12}f_4)\times(e^{\kappa x}-1)b_{22}.$$

It remains to calculate f_1, f_4:

$$\begin{aligned}
f_1=&-\frac{k_1^2}{4\lambda_c}[a_1^2M(c_{11}-\alpha c_{21})+2a_1a_3M(c_{12}-\alpha c_{22})\\
&+a_3^2M(d_{12}-\alpha d_{22})],\\
f_4=&-\frac{k_1^2}{4\lambda_c}\{a_1^2[c_{21}(\lambda_c+\alpha^2M)-c_{11}\alpha M]\\
&+2a_1a_3[c_{22}(\lambda_c+\alpha^2M)-c_{12}\alpha M]\\
&+a_3^2[d_{22}(\lambda_c+\alpha^2M)-d_{12}\alpha M]\}.
\end{aligned}$$

For the coefficients of v_2^2 we write down $b_2(x)$ and $b_5(x)$

$$b_2(x)=\frac{1}{\kappa}(b_{22}f_2-b_{12}f_5)xb_{12}+\frac{1}{\kappa^2}(b_{11}f_2+b_{12}f_5)\times(e^{\kappa x}-1)b_{12},$$

$$b_5(x)=-\frac{1}{\kappa}(b_{22}f_2-b_{12}f_5)xb_{11}+\frac{1}{\kappa^2}(b_{11}f_2+b_{12}f_5)\times(e^{\kappa x}-1)b_{22}.$$

The nondegenerate case involving the calculations of the coefficients of v_1v_2 is done similarly.

IV. DISCUSSION

The objective of the present analysis is to establish a correlation between physical parameters of the medium and measurable nonlinear acoustic parameters.

TABLE I. Physical properties of the medium.

Bulk modulus of grains	K_r (Pa)	3.6×10^{10}
Bulk modulus of fluid/gas	K_f (Pa)	2.4×10^{9}
Density of grains	ρ (kg/m^3)	2.65×10^{3}
Density of fluid/gas	ρ_f (kg/m^3)	1.0×10^{3}
Shear coefficient	μ_0 (Pa)	1.0×10^{8}
Posority	β	0.4
Added mass coefficient	a	0.25
Poisson's ratio	ν	0.33

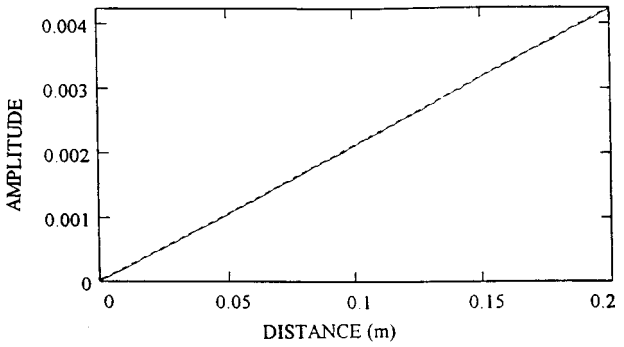


FIG. 3. Amplitude of the second harmonic versus distance of propagation for a fluid-saturated medium. The solid line is for the full amplitude $|V_N|$, the dashed line is for the slow wave contribution (amplitude $|b_1|$).

A. Porous media parameters

The parameters or coefficients used in Biot's theory can be divided into three major categories:

(1) Structural parameters:
Density ρ, bulk K_r, and shear μ_r moduli of grains.
Density ρ_f and bulk modulus K_f of fluid in pores.
Bulk K_s and shear μ moduli of skeleton frame.
Porosity β.
Nonlinear coefficients D, F, and G.

(2) Dissipation parameters:
Fluid viscosity η.
Permeability k.
Frame loss coefficients.

(3) More dependent parameters:
Biot's coefficients α and M.
Darcy's coefficient m.

In the present analysis we will focus on the structural parameters. In this category the porosity is the governing parameter for porous media. For example, some structural parameters, such as frame bulk and shear moduli show experimental dependence on the porosity. The nonlinear coefficients must also depend on porosity since numerous experiments show that they are two to three orders of magnitude greater in porous media than in nonporous solids or fluids.[8] Moreover, when $\beta=0$ (no pores), K_s must be equal to K_r and $\mu=\mu_r$. On the other hand, if the porosity is close to 1 (fluid), K_s should be close to K_f and μ goes to zero. The difference in behavior at the two limits requires a noticeable dependence on the porosity.

An empirical relationship between shear modulus and porosity for porous sediments

$$\mu(\beta)=\mu_0\left(\frac{1-\beta}{\beta}\right)^{1.12} \tag{33}$$

is used in Ref. 18. Here μ_0 is an experimentally defined constant for the particular medium. According to this formula, as the porosity β tends to one (pure fluid), the shear modulus approaches zero, as it should. However, as β tends to zero (the porous medium is approaching that of a solid), the shear modulus approaches infinity. We modify the empirical formula to read

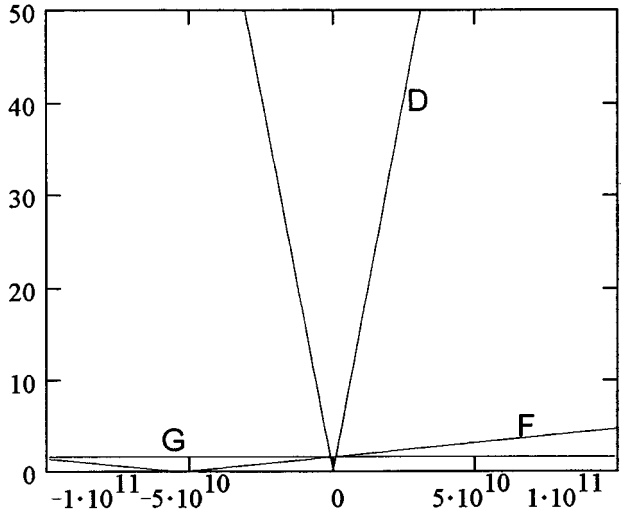


FIG. 4. Dependence of the effective nonlinear parameter Γ on the parameters D, F, and G.

$$\mu(\beta)=\mu_0\left(\frac{1-\beta}{\beta+d}\right)^{1.12}, \tag{34}$$

where d is a constant chosen so that the value of the shear modulus of the frame μ coincides with the shear modulus of the grains for $\beta=0$; $\mu(0)=\mu_r$. An example of shear modulus dependence on porosity for constants $\mu_0=10^8$, $d=0.0123$, corresponding to the empirical relationship (33) and its modification (34) is shown in Fig. 1.

The bulk modulus of the frame also becomes porosity dependent

$$K_s(\beta)=\mu(\beta)\left(\frac{2\nu}{1-2\nu}+\frac{2}{3}\right), \tag{35}$$

where ν is Poisson's ratio (we are considering $\nu=$constant since Poisson's ratio generally has a very limited range of variation).

The Biot coefficients α and M become functions of the porosity as well,

$$\begin{aligned}\alpha(\beta)&=1-\frac{K_s(\beta)}{K_r},\\ M(\beta)&=K_r^2\Big/\left\{K_r\left[1+\beta\left(\frac{K_r}{K_f}-1\right)\right]-K_s(\beta)\right\}.\end{aligned} \tag{36}$$

With this modification to the Biot coefficients (β modification), the sound speeds of the fast wave and the slow wave in a fluid-saturated medium assume the appropriate values as the porosity changes its values on the interval [0,1]; this is depicted in Fig. 2. In particular, notice that the fast wave approaches the value of the sound speed in a fluid while the speed of the slow wave approaches zero (disappears) as β tends to unity.

B. Second-harmonic generation

Consider numerical examples of the second-harmonic generation versus distance of propagation using the solution derived in Sec. III. The calculations were carried out for

TABLE II. Nonlinear parameters of some media.

$\beta=0$	Porous media $0<\beta<1$				$\beta=1$
Nonporous solid media	Marble	Sandstone and soils	Marine sediments	Foam plastic	Fluids and gases
$\Gamma<15$	$\Gamma=800$	$\Gamma=10^3-10^4$	$\Gamma=10^2-10^3$	$\Gamma=200$	$\Gamma<10$

nondissipative porous fluid-saturated sediments similar to Stoll and Kan's examples,[19] Table I, for equal initial amplitudes for the slow and fast waves.

As can be seen from (29), there are three components contributing to the nonlinear solution, the slow wave v_1^2, the fast wave v_2^2, and their product $v_1 v_2$. The results of calculations shown in Fig. 3 indicate that in a water-saturated medium, the slow wave is the main contributor to the second harmonic. These results also demonstrate quasilinear growth of the second harmonic with distance, which agrees with experimental observations.[6]

C. Nonlinear parameters

There are three nonlinear parameters in our model, D, G, and F. However, in practice, it is more convenient to use a parameter which is proportional to the ratio of the second-harmonic amplitude to the first-harmonic amplitude. This parameter is commonly used to describe the nonlinearity of liquid[20] and solid elastic media,[21] and adopted as a practical measure of nonlinearity for poroelastic media.[1,2,6]

The one-dimensional equation of motion for nonporous nondissipative elastic media can be written, to the first order in the nonlinearity, in the form[6]

$$\frac{\partial^2}{\partial x^2}u(x,t)-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}u(x,t)=-\Gamma\frac{\partial}{\partial x}\left[\frac{\partial}{\partial x}u(x,t)\right]^2, \tag{37}$$

where Γ is the nonlinear parameter, $u(x,t)$ is the particle displacement, and c is the compressional velocity. The perturbation technique produces, for an initial harmonic wave of frequency ω and constant amplitude A, the solution

$$u(x,t)=A\exp(i(kx-\omega t))+\tfrac{1}{2}\Gamma A^2k^2x \times\exp(2i(kx-\omega t)), \tag{38}$$

where $k=\omega/c$ is the wave number.

Comparing solution (38) with the our solution

$$u(x,t)=A\exp(i(\omega t-k_{1,2}x))+|u_N(x)|\exp(2i\omega t), \tag{39}$$

where the second-harmonic amplitude $u_N(x)$ was defined in Sec. III, we can introduce an effective nonlinear parameter for the second-harmonic generation

$$\Gamma(\beta,\mu,K_s,D,F,G)=\frac{2}{k^2A^2}\frac{|u_N(x)|}{x}. \tag{40}$$

The parameter is practically independent of x since the second-harmonic amplitude $u_N(x)$ is a quasilinear function of x as shown in the numerical example. The choice of wave number k_1 or k_2 in formula (40) is determined by the major contributor to the second harmonic, therefore, k_2 has to be used as implied from the demonstrated results.

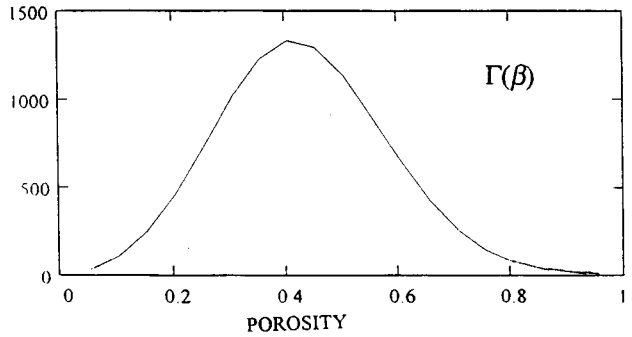


FIG. 5. Dependence of the effective nonlinear parameter Γ on the porosity β.

Numerical investigations reveal an important result. The effective nonlinear parameter is proportional to the nonlinear parameter D and is almost independent of F and G as shown in Fig. 4.

The nonlinear parameters of the porous media depend on porosity. The available experimental data[1–8,10,20] summarized in Table II indicates that the nonlinear coefficient Γ is two to three orders of magnitude greater in porous media than in nonporous solid or fluid media. This means that if we perform a virtual experiment measuring the nonlinear coefficient Γ as a function of porosity β for the same medium, then the measured dependence is likely to be such as shown in Fig. 5. Since $\Gamma\propto D$, then it follows that $D=D(\beta)$. This statement can be tested by introducing the dependence $D(\beta)$ (Fig. 6) to reproduce $\Gamma(\beta)$ shown in Fig. 5.

This implication of dependence of nonlinear parameters on porosity may have very important practical utilization. Since the developed model establishes a correlation between the measurable effective nonlinear parameter and the structural parameters of the porous media, the nonlinear measurements can be used as a diagnostic tool for the media.

This can be a very sensitive tool because porosity induces orders of magnitude change for the nonlinear coefficients and only a few percent change for the linear parameters.

V. SUMMARY

In this study, we derive nonlinear acoustic wave equations for a poroelastic medium. The equations are based on

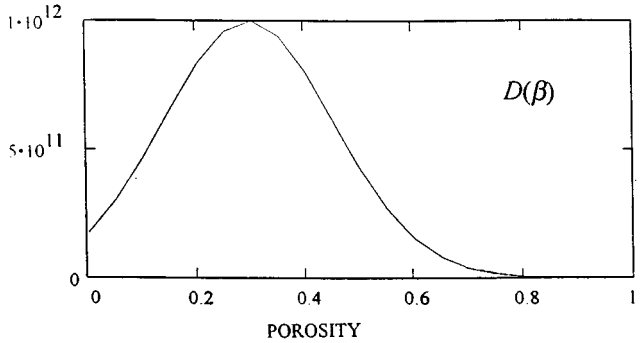


FIG. 6. Dependence of the nonlinear parameter D on the porosity β.

the semilinear approximation of Biot's poroelasticity theory. The slowly varying amplitude assumption and a perturbation technique are used to solve the equations. The contributions of the slow and fast longitudinal waves in the nonlinear solution are analyzed for fluid saturated media.

The dependence of the linear and nonlinear structural parameters on porosity is empirically introduced.

The developed model establishes a correlation between the measurable effective nonlinear parameter and structural parameters of the porous medium. This suggests employing nonlinear measurements as a diagnostic tool for porous media.

ACKNOWLEDGMENTS

The work of the first author was supported in part by NASA Grant No. NAGW-4586.

[1] V. N. Bakunin and A. G. Protosenya, ''Nonlinear effects in travel of Elastic waves through rocks,'' Trans. (Dokl.) USSR Acad. Sci. Earth Sci. Sec. **263**, 314–316 (1981).

[2] P. A. Johnson, T. J. Shankland, R. J. O'Connell, and J. N. Albright, ''Nonlinear generation of elastic waves in crystalline rock,'' J. Geophys. Res. **92**, 3597–3602 (1987).

[3] P. A. Johnson and T. J. Shankland, ''Nonlinear generation of elastic waves in granite and sandstone: Continuous wave and travel time observations,'' J. Geophys. Res. **94**, 17729–17733 (1989).

[4] I. A. Beresnev and A. V. Nikolaev, ''Experimental investigations of nonlinear seismic effects,'' Phys. Earth Planetary Int. **50**, 83–87 (1988).

[5] I. Yu. Belyaeva and E. M. Timanin, ''Experimental study of the nonlinear properties of porous elastic media,'' Sov. Phys. Acoust. **37**, 533–534 (1991).

[6] G. D. Meegan, P. A. Johnson, R. A. Guyer, and K. R. McCall, ''Observations of nonlinear elastic wave behavior in sandstone,'' J. Acoust. Soc. Am. **94**, 3387–3391 (1993).

[7] K. W. Winkler, ''Measurements of third-order elastic constants in rocks,'' J. Acoust. Soc. Am. **98**, 2886 (1995).

[8] L. A. Ostrovsky, ''Wave processes in media with strong acoustic nonlinearity,'' J. Acoust. Soc. Am. **90**, 3332–3337 (1991).

[9] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity* (Pergamon, New York, 1986).

[10] K. V. D. Abeele, P. A. Johnson, and J. A. TenCate, ''Elastic nonlinearity in rock: On the relative importance between higher-order elastic constants and hysteresis,'' J. Acoust. Soc. Am. **98**, 2887 (1995).

[11] A. N. Norris and M. A. Grinfeld, ''Nonlinear poroelasticity for a layered medium,'' J. Acoust. Soc. Am. **98**, 1138–1146 (1995).

[12] D. M. Donskoy and A. M. Sutin, ''Nonlinear scattering and propagation of longitudinal acoustic waves in porous media,'' Sov. Phys. Acoust. **30**, 358–361 (1984).

[13] L. A. Ostrovsky, ''Nonlinear properties of an elastic medium with cylindrical pores,'' Sov. Phys. Acoust. **35**, 286–289 (1989).

[14] *Acoustics, Elasticity, and Thermodynamics of Porous Media. Twenty-One Papers by M. A. Biot* (Acoustical Society of America, Woodbury, NY, 1992).

[15] M. A. Biot, ''Nonlinear and semilinear rheology of porous solids,'' J. Geophys. Res. **78**, 4924–4937 (1973).

[16] M. A. Grinfeld and A. N. Norris, ''Acoustoelasticity theory and applications for fluid-saturated porous media,'' J. Acoust. Soc. Am. **100**, 1368–1374 (1996).

[17] A. N. Norris, ''Stoneley wave attenuation and dispersion in permeable formations,'' Geophysics **54**, 330–341 (1989).

[18] T. Yamamoto, M. V. Trevorrow, M. Badiey, and A. Turgut, ''Seabed porosity and shear modulus inversion using surface gravity (water) wave-induced seabed motion,'' Geophys. J. Int. **98**, 173–182 (1989).

[19] R. D. Stoll and T. K. Kan, ''Reflection of acoustic waves at a water-sediment interface,'' J. Acoust. Soc. Am. **70**, 149–156 (1981).

[20] R. T. Beyer, ''Parameter of nonlinearity in fluids,'' J. Acoust. Soc. Am. **32**, 719–721 (1960).

[21] M. A. Breazeale and J. Philip, ''Determination of third-order elastic constants from ultrasonic harmonic generation measurements,'' in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1984), Vol. XVII.

On Rayleigh wave nonlinearity, and analytical approximation of the shock formation distance

Ekaterina Yu. Knight, Mark F. Hamilton, Yurii A. Il'inskii, and Evgenia A. Zabolotskaya
Department of Mechanical Engineering, The University of Texas at Austin, Austin, Texas 78712-1063

(Received 30 July 1996; accepted for publication 31 January 1997)

Rayleigh wave nonlinearity is investigated theoretically. First, spectral forms of the evolution equations for Rayleigh waves and sound waves are used to compare the relative efficiencies of sum and difference frequency generation. Next, time domain forms of the same evolution equations are used to compare the relative importance of local and nonlocal nonlinearity on the distortion of Rayleigh waves in the preshock region. Finally, new analytical approximations are derived for the shock formation distance and the corresponding coefficient of nonlinearity for Rayleigh waves. © *1997 Acoustical Society of America.* [S0001-4966(97)06010-4]

PACS numbers: 43.25.Fe [MAB]

INTRODUCTION

It is known from previous work[1,2] that solutions for the horizontal velocity component in nonlinear Rayleigh waves exhibit similarities with solutions for simple waves in fluids.[3,4] Phenomena such as shock formation and acoustic saturation can be observed in Rayleigh waves as well as in sound waves. However, several important features of the behavior of Rayleigh waves cannot be explained on the basis of nonlinearity associated with simple waves.[5,6] For example, cusping of the waveform at the shock[1,2] is a feature that is unique to nonlinear Rayleigh waves. To take advantage of existing knowledge about the distortion of plane sound waves, and to understand better the distinctive features of nonlinear Rayleigh waves, we shall make direct comparisons of solutions obtained for these two cases.

Spectral forms of the evolution equations for Rayleigh waves and sound waves are compared in Sec. I in order to identify a coefficient of nonlinearity for Rayleigh waves. The frequency dependence of this coefficient is examined. In Sec. II, the time domain form of the evolution equation is used to separate local and nonlocal nonlinear effects in Rayleigh waves, the relative importance of which in the preshock region is investigated in Sec. III. Conclusions drawn from both the time and frequency domain analyses motivate the derivation in Sec. III of a new approximate analytical expression for the Rayleigh wave shock formation distance, Eq. (33) below, and for an effective constant coefficient of nonlinearity associated with this distance, Eq. (32). As for sound waves in fluids, the shock formation distance is the length scale that characterizes the rate of nonlinear distortion. Predictions based on these analytical approximations are compared with direct numerical calculations for ten different materials, and agreement within several percent is obtained.

I. FREQUENCY DEPENDENCE OF THE COEFFICIENT OF NONLINEARITY

We begin with the spectral evolution equation derived previously[7] for nonlinear Rayleigh waves in lossless isotropic solids:

$$\frac{\partial \tilde{v}}{\partial x} = -\frac{\omega^2 \mu}{\rho c_R^4 \zeta} \int_{-\infty}^{\infty} \frac{\omega'(\omega-\omega')}{|\omega'(\omega-\omega')|} R(\omega', \omega-\omega') \times \tilde{v}(x,\omega')\tilde{v}(x,\omega-\omega') \frac{d\omega'}{2\pi}, \tag{1}$$

where $\tilde{v}(x,\omega)$ is a spectral amplitude related to particle velocity, μ is the shear modulus, ρ is the mass density, c_R is the small-signal Rayleigh wave speed, and the following dimensionless quantities are based on the small-signal speeds of the longitudinal (c_l) and shear (c_t) waves:

$$\zeta = \xi_t + \xi_t^{-1} + (\xi_l + \xi_l^{-1})\eta^2 + 4\eta,$$

$$\xi_t = (1-\xi^2)^{1/2}, \quad \xi_l = (1-\xi^2 c_t^2/c_l^2)^{1/2},$$

$$\eta = -2\xi_t/(1+\xi_t^2), \quad \xi = c_R/c_t,$$

for which $\xi_t + \eta < 0$ and $1 + \eta\xi_l > 0$. The continuous form of the nonlinearity matrix used above, $R(\omega_1, \omega_2)$, is obtained from the discrete form used in previous work,[2] R_{lm}, by making the following substitutions in the latter: $l \to \omega_1$ and $m \to \omega_2$. For the purpose of comparison below with the spectral form of the evolution equation for sound waves, it is convenient to rewrite Eq. (1) in terms of spectral amplitudes of the horizontal (in-plane) velocity component at the free surface of the elastic half-space. The horizontal velocity spectrum at the interface is given by

$$\tilde{v}_x(x,\omega) = i\frac{\omega}{|\omega|}(\xi_t + \eta)\tilde{v}(x,\omega), \tag{2}$$

in terms of which Eq. (1) becomes

$$\frac{\partial \tilde{v}_x}{\partial x} = -\frac{i\omega\mu}{\rho c_R^4 \zeta |\xi_t + \eta|} \int_{-\infty}^{\infty} F(\omega', \omega-\omega') \times \tilde{v}_x(x,\omega')\tilde{v}_x(x,\omega-\omega') \frac{d\omega'}{2\pi}, \tag{3}$$

where the notation[5]

$$F(\omega_1,\omega_2)=|\omega_1+\omega_2|R(\omega_1,\omega_2) \tag{4}$$

is employed. Fourier transformation yields the horizontal velocity waveform:

$$v_x(x,\tau)=\int_{-\infty}^{\infty}\tilde{v}_x(x,\omega)e^{-i\omega\tau}\,\frac{d\omega}{2\pi}, \tag{5}$$

where $\tau=t-x/c_{\mathrm{R}}$ is a retarded time.

Now consider the nonlinear evolution equation for plane progressive sound waves in a lossless medium,[4]

$$\frac{\partial v_l}{\partial x}=\frac{\beta_l}{c_l^2}\,v_l\,\frac{\partial v_l}{\partial t'}, \tag{6}$$

where v_l is the particle velocity in the longitudinal wave, $t'=t-x/c_l$ is the retarded time, and β_l is the coefficient of nonlinearity defined by[8]

$$\beta_l=(\gamma+1)/2, \quad \text{gas}, \tag{7}$$

$$=1+B/2A, \quad \text{liquid}, \tag{8}$$

$$=-\left(\frac{3}{2}+\frac{\mathscr{A}+3\mathscr{B}+\mathscr{C}}{\rho c_l^2}\right), \quad \text{solid}. \tag{9}$$

Here, γ is the ratio of specific heats, B/A is the nonlinearity parameter in the equation of state for an arbitrary liquid,[9] and $\mathscr{A}$, $\mathscr{B}$, and $\mathscr{C}$ are Landau's[10] third-order elastic constants [Eq. (9) may be written in terms of Murnaghan's[11] third-order constants l, m, and n by replacing $\mathscr{A}+3\mathscr{B}+\mathscr{C}$ with $l+2m$]. To obtain the spectral form of the evolution equation for sound waves, express the particle velocity as

$$v_l(x,t')=\int_{-\infty}^{\infty}\tilde{v}_l(x,\omega)e^{-i\omega t'}\,\frac{d\omega}{2\pi}. \tag{10}$$

Fourier transformation of Eq. (6) thus yields[12]

$$\frac{\partial\tilde{v}_l}{\partial x}=-\frac{i\omega\beta_l}{2c_l^2}\int_{-\infty}^{\infty}\tilde{v}_l(x,\omega')\tilde{v}_l(x,\omega-\omega')\,\frac{d\omega'}{2\pi}. \tag{11}$$

Via comparison of Eqs. (3) and (11), we identify the following coefficient of nonlinearity for the interaction of Rayleigh waves with frequencies ω_1 and ω_2:

$$\beta_{\mathrm{R}}(\omega_1,\omega_2)=\frac{2\mu F(\omega_1,\omega_2)}{\rho c_{\mathrm{R}}^2\zeta|\xi_t+\eta|}\,. \tag{12}$$

Equation (3) may now be rewritten as

$$\frac{\partial\tilde{v}_x}{\partial x}=-\frac{i\omega}{2c_{\mathrm{R}}^2}\int_{-\infty}^{\infty}\beta_{\mathrm{R}}(\omega',\omega-\omega')\times\tilde{v}_x(x,\omega')\tilde{v}_x(x,\omega-\omega')\,\frac{d\omega'}{2\pi} \tag{13}$$

to facilitate comparison with Eq. (11). Note that Eqs. (11) and (13) are formally equivalent only if the kernel $\beta_{\mathrm{R}}(\omega_1,\omega_2)$ for Rayleigh waves does not depend on frequency. It is specifically the frequency dependence through $F(\omega_1,\omega_2)$ that corresponds to the nonlocal character of nonlinear Rayleigh wave interactions and distinguishes their behavior from sound waves.[5]

We therefore proceed to investigate the function $F(\omega_1,\omega_2)$, the expression for which is[5]

$$\begin{aligned}F(\omega_1\omega_2)=&\frac{\alpha'|\omega_1+\omega_2|}{|\omega_1+\omega_2|\xi_l+|\omega_1|\xi_t+|\omega_2|\xi_t}+\frac{\alpha'|\omega_1+\omega_2|}{|\omega_1+\omega_2|\xi_t+|\omega_1|\xi_l+|\omega_2|\xi_t}+\frac{\alpha'|\omega_1+\omega_2|}{|\omega_1+\omega_2|\xi_t+|\omega_1|\xi_t+|\omega_2|\xi_l}\\&+\frac{\beta'|\omega_1+\omega_2|}{|\omega_1+\omega_2|\xi_t+|\omega_1|\xi_l+|\omega_2|\xi_l}+\frac{\beta'|\omega_1+\omega_2|}{|\omega_1+\omega_2|\xi_l+|\omega_1|\xi_t+|\omega_2|\xi_l}+\frac{\beta'|\omega_1+\omega_2|}{|\omega_1+\omega_2|\xi_l+|\omega_1|\xi_l+|\omega_2|\xi_t}\\&+\frac{3\gamma'|\omega_1+\omega_2|}{|\omega_1+\omega_2|\xi_l+|\omega_1|\xi_l+|\omega_2|\xi_l},\end{aligned} \tag{14}$$

where the constants α', β', and γ' are defined in terms of Landau's notation and alternatively in terms of Murnaghan's notation in Appendix A of Ref. 7. Examination of $F(\omega_1,\omega_2)$ reveals that it depends only on the ratio ω_1/ω_2, as opposed to ω_1 and ω_2 independently, and it can therefore be written as a function of a single argument:

$$F(\omega_1,\omega_2)=f(\omega_1/\omega_2). \tag{15}$$

The explicit dependence on ω_1/ω_2 motivates the introduction of polar coordinates:

$$F(\omega_1,\omega_2)=f_\phi(\phi), \quad \tan\phi=\omega_1/\omega_2. \tag{16}$$

To illustrate the main features of β_{R}, in Fig. 1 we plot $f(\omega_1/\omega_2)$ and $f_\phi(\phi)$ for steel (using material parameters cited previously[1,2]). Positive values of ω_1/ω_2 $(0\leqslant\phi\leqslant\pi/2)$ correspond to sum frequency generation, negative values to difference frequency generation. The first important feature to note is that β_{R} varies slowly with ω_1/ω_2 for sum frequency generation $(\omega_1/\omega_2>0)$ and rapidly for difference frequency generation $(\omega_1/\omega_2<0)$. For example, β_{R} vanishes at the point $\omega_1/\omega_2=-1$, which corresponds to generation of a signal at zero frequency. Second, it can be seen that in comparison to sum frequency generation, difference frequency generation is less efficient in Rayleigh waves than in sound waves, because the coefficient of nonlinearity for the latter is a constant.

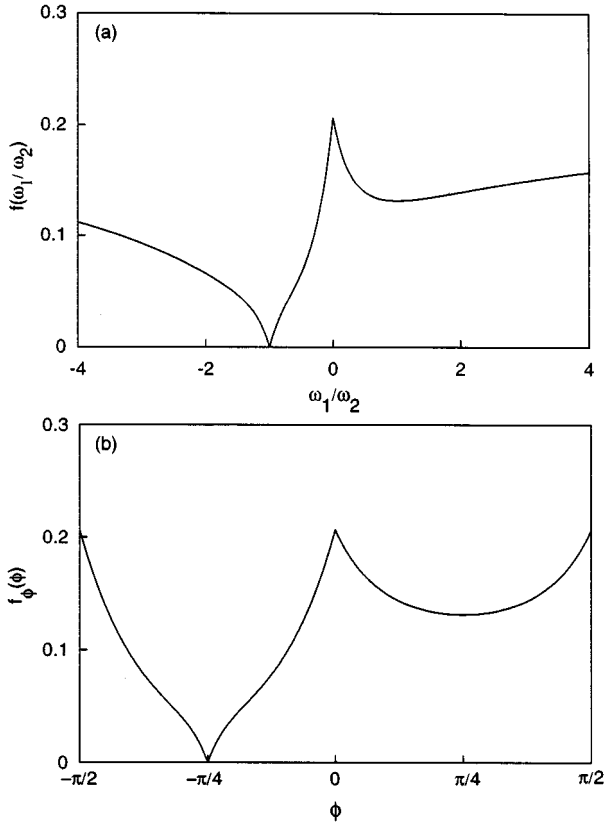


FIG. 1. Frequency dependence of the kernel $F(\omega_1,\omega_2)$, plotted (a) as a function of ω_1/ω_2 and (b) as a function of the polar angle $\phi = \arctan(\omega_1/\omega_2)$.

II. TIME DOMAIN EVOLUTION EQUATION

To develop an analytical approximation of the shock formation distance, we turn now to the time domain form of Eq. (3),[5]

$$\frac{\partial v_x}{\partial x} = \frac{\overline{\beta_R}}{c_R^2} v_x \frac{\partial v_x}{\partial \tau} + \frac{C}{c_R^2}\frac{\partial}{\partial \tau}\int\int_T \frac{B(\tau_1/\tau_2)}{\tau_1^2+\tau_2^2} \times v_x(x,\tau-\tau_1)v_x(x,\tau-\tau_2)d\tau_1 d\tau_2, \tag{17}$$

where T is defined as either the duration of a pulse or the period of a periodic waveform, and we identify

$$\overline{\beta_R} = \frac{2\mu\overline{f}}{\rho c_R^2 \zeta|\xi_t+\eta|} \tag{18}$$

as a coefficient of *local* nonlinearity, where

$$\overline{f} = \frac{1}{\pi}\int_0^\pi f_\phi(\phi)d\phi. \tag{19}$$

The following definitions are provided in Ref. 5: the function B in Sec. III B, the constant C in Appendix A, and the constant f in Appendix C. Attention below is focused mainly on the first term on the right-hand side of Eq. (17), which accounts for all local nonlinearity in Rayleigh waves.

The first term on the right-hand side corresponds to the rate of waveform distortion at an instant τ determined by the value of v_x at that same instant. This term is therefore associated with local nonlinearity. The second term contains integrals over the entire waveform, i.e., the distortion at a particular instant on the waveform depends on values of v_x at all instants τ. The second term is thus associated with nonlocal nonlinearity. Further discussion of differences between local and nonlocal nonlinearity is given in Ref. 5.

Note that if we set $B=0$, Eq. (17) becomes formally equivalent to Eq. (6) for finite-amplitude compression waves. Moreover, Eq. (18) is very similar to Eq. (12), with β_R obtained by averaging $\beta_R(\omega_1,\omega_2)$ over all possible interaction frequencies. However, whereas the function β_R accounts for both local and nonlocal nonlinearity, the constant β_R is associated exclusively with local nonlinear effects. The relative contributions of local and nonlocal nonlinearity to the distortion of a Rayleigh wave in the preshock region are estimated in Sec. III by comparing two shock formation distances, one calculated with only local nonlinearity taken into account, the other with both local and nonlocal nonlinearity.

III. SHOCK FORMATION DISTANCE

The shock formation distance is an important characteristic of acoustical media because it represents the appropriate length scale for nonlinear processes. The shock formation distance for a plane finite-amplitude sound wave with source condition

$$v_l(0,t) = v_{l0}\sin\omega_0 t, \tag{20}$$

radiated into a lossless medium, is[4,8]

$$\overline{x_l} = \frac{1}{|\beta_l|\epsilon_l k_l}, \tag{21}$$

where v_{l0} (assumed positive) is the peak particle velocity, $\epsilon_l = v_{l0}/c_l$ is the corresponding acoustic Mach number (or equivalently, peak strain), and $k_l = \omega_0/c_l$ is the wave number at the source. This result follows directly from Eq. (6) for the location where the signal first develops a vertical tangent to its waveform. Because of the complicated form of Eq. (17), there is no corresponding exact analytical expression at the present time for a Rayleigh wave shock-formation distance.

It is the horizontal velocity component in a Rayleigh wave that bears most resemblence to acoustic waveforms, because a sinusoid at the source distorts during propagation into a sawtoothlike profile.[1,2] In contrast, the vertical velocity waveform, which at the interface is proportional to the Hilbert transform of the horizontal velocity waveform, exhibits spikes where shocks appear in the latter. We consider here the following plane-wave source condition for a Rayleigh wave:

$$v_x(0,t) = v_{x0}\sin\omega_0 t, \tag{22}$$

$$v_z(0,t) = -v_{z0}\cos\omega_0 t, \tag{23}$$

where v_{x0} and v_{z0} are the amplitudes of the horizontal and vertical components, respectively, each assumed positive and related by

$$v_{x0} = \left(\frac{|\xi_t+\eta|}{1+\eta\xi_l}\right)v_{z0}. \tag{24}$$

We have taken positive z to be directed outward from the solid. For theoretical purposes it is convenient to exploit the

analogy with sound waves and thus use the velocity v_{x0} as the reference source amplitude, noting that v_{z0} is often the more easily measured quantity in experiments. We thus define a Rayleigh wave shock-formation distance $\overline{x_R}$ that corresponds to the source condition in Eq. (22), together with an associated coefficient of nonlinearity β_x, as follows:

$$\overline{x_R}=\frac{1}{|\beta_x|\epsilon_x k_0}, \tag{25}$$

where $\epsilon_x=v_{x0}/c_R$ is the peak horizontal Mach number (or equivalently, peak horizontal strain) and $k_0=\omega_0/c_R$ is the wave number at the source. Our objective below is to develop a simple analytical approximation of β_x for arbitrary isotropic solids, such that an accurate estimate of the shock formation distance is readily calculated given the amplitude (ϵ_x) and frequency (k_0) at the source. In terms of the peak amplitude of the vertical component at the source, we may use Eq. (24) to rewrite Eq. (25) as

$$\overline{x_R}=\left(\frac{1+\eta\xi_l}{|\xi_t+\eta|}\right)\frac{1}{|\beta_x|\epsilon_z k_0}, \tag{26}$$

where $\epsilon_z=v_{z0}/c_R$.

Candidate expressions approximating the shock formation distance for Rayleigh waves have been proposed in previous articles. In Ref. 1, the quantity defined (in our notation here) by

$$x_0=\frac{2\rho c_R^2\zeta|\xi_t+\eta|}{\mu\epsilon_x k_0} \tag{27}$$

appeared as a natural length scale for rendering the spectral evolution equations derived there in dimensionless form. Moreover, the numerical calculations reported in that same article for Rayleigh wave propagation in steel reveal that shock formation occurs in the region $x_0<x<2x_0$, and therefore x_0 in this particular case is an appropriate length scale for characterizing the nonlinearity.

In a subsequent article,[2] it was noted that x_0 is not an appropriate nonlinear length scale for all materials because it does not depend on third-order elastic constants. An alternative scale, which does depend on third-order constants, was inferred by comparing quasilinear solutions for second harmonic generation in a Rayleigh wave and in a sound wave. The approximate shock formation distance proposed by those authors is $\overline{x_{11}}=x_0/8|R_{11}|$, rewritten here as

$$\overline{x_{11}}=\frac{1}{|\beta_{11}|\epsilon_x k_0}, \tag{28}$$

where

$$\beta_{11}=\frac{4\mu R_{11}}{\rho c_R^2\zeta|\xi_t+\eta|} \tag{29}$$

is the corresponding coefficient of nonlinearity, and R_{11}, defined in Eq. (7) of Ref. 2, is the element of the nonlinearity matrix R_{lm} associated with second harmonic generation. Equation (28) yields $\overline{x_{11}}=1.9x_0$ for steel, which is in nominal agreement with the predicted waveform evolution,[1,2] but the authors did not assess the accuracy of $\overline{x_{11}}$ via precise comparison with numerical calculations of waveform distortion.

TABLE I. Values of the analytical estimates $\overline{x_{11}}$ and $\overline{x_+}$ expressed in terms of the numerical calculation of $\overline{x_R}$, and the values of β_+, for different materials. Values for the third-order elastic constants for the first five materials were taken from Ref. 13, and from Ref. 14 for the second five.

	Analytical results		
Material	$\overline{x_{11}}/\overline{x_R}$	$\overline{x_+}/\overline{x_R}$	β_+
Polystyrene	1.20	1.01	0.84
Nickel steel 535	1.18	1.01	0.85
Steel 60N2S2A	1.17	1.01	1.03
Armco iron	1.15	1.01	2.03
Pyrex	1.10	1.04	−2.42
Barre granite	1.09	1.04	200
Dry D82 marble	1.12	1.03	593
Water-saturated F32, thermally cracked	1.08	1.04	1523
Dry F32 Fountainebleau sandstone	1.11	1.03	3201
Dry F32, thermally cracked	1.11	1.03	4192

Our method for calculating the shock formation distance numerically is described in Appendix A. In brief, it consists of solving coupled equations for the spectral amplitudes [namely, the discrete form of Eq. (1)], with the source condition prescribed by Eqs. (22)–(24). The derivative of the horizontal velocity waveform is monitored as a function of distance, and the location where it becomes infinite is identified as the numerical approximation of the shock formation distance $\overline{x_R}$. Values of $\overline{x_R}$ obtained in this way are taken hereafter as the basis of comparison for the proposed analytic approximations of the shock formation distance.

Calculations were performed for the ten materials listed in the first column of Table I. These particular materials were chosen for two reasons. First, they are characterized as isotropic solids by five elastic constants, two at second order (e.g., bulk and shear moduli) and three at third order, numerical values for which are reported in the literature (the third-order elastic constants were obtained from Refs. 13 and 14). Second, the materials encompass a wide range of nonlinearity (note that the values of $\beta_+\simeq\beta_x$ in the last column, discussed below, span more than three orders of magnitude).

We now return to analytical approximations of the shock formation distance. In the second column of Table I we assess the accuracy of Eq. (28) by calculating the ratio $\overline{x_{11}}/\overline{x_R}$. It is seen that Eq. (28) overestimates numerical calculations of the shock formation distances for the given materials by 8%–20%.

As noted above, the analytical prediction of the shock formation distance given by $\overline{x_{11}}$ takes only second harmonic generation into account. We now develop an improved analytical estimate that accounts for other spectral interactions. As a first step, consider the shock formation distance that is

predicted by Eq. (17) if all nonlocal nonlinearity is ignored. Thus set $B=0$, in which case the shock formation distance is given by Eq. (25) with β_x replaced by β_{R}, and define

$$\overline{x_L}=\frac{1}{|\overline{\beta_{\mathrm{R}}}|\epsilon_x k_0}. \tag{30}$$

Calculations reveal that $\overline{x_L}$ overestimates $\overline{x_{\mathrm{R}}}$ by 27%–34% for the materials in Table I (we do not list these results here; see Ref. 15), i.e., the predictions exceed those of the estimate provided by $\overline{x_{11}}$. On the other hand, the fact that $\overline{x_L}$ approximates $\overline{x_{\mathrm{R}}}$ to within about 30% demonstrates that local nonlinearity dominates nonlocal nonlinearity in the early stages of waveform distortion.

The results presented above may be used to develop a better analytical estimate for the shock formation distance. Specifically, we continue to focus on local nonlinearity as the principal contribution to waveform distortion up to the point of shock formation. Recall from Eq. (18) that β_{R} in Eq. (30) is proportional to the average of f_ϕ over all interaction frequencies, where f_ϕ characterizes the strength of individual spectral interactions. In the preshock region, sum frequencies are generated more efficiently than difference frequencies, and it is sum frequency generation that is primarily responsible for shock formation. From Fig. 1, we see that averaging f_ϕ over all interaction frequencies, including the small values of f_ϕ associated with difference frequency generation, yields an effective nonlinearity coefficient β_{R} that is too low in comparison with the nominal interaction strength associated with sum frequency generation. For this reason, $\overline{x_L}$ substantially overestimates $\overline{x_{\mathrm{R}}}$. It therefore seems reasonable to estimate the effective nonlinearity in the early stages of waveform distortion by averaging f_ϕ over the domain $0 \leqslant \phi \leqslant \pi/2$ (i.e., excluding difference frequency generation), and calculating the corresponding coefficient of nonlinearity and shock formation distance accordingly.

Thus, in place of Eq. (19) define

$$\overline{f_+}=\frac{2}{\pi}\int_0^{\pi/2} f_\phi(\phi)\,d\phi, \tag{31}$$

where an analytical expression for $\overline{f_+}$ is given by Eq. (B3). Similarly, in place of Eq. (18) define the coefficient of nonlinearity

$$\beta_+=\frac{2\mu\overline{f_+}}{\rho c_{\mathrm{R}}^2\zeta|\xi_t+\eta|}, \tag{32}$$

and in place of Eq. (30) define the corresponding shock formation distance

$$\overline{x_+}=\frac{1}{|\beta_+|\epsilon_x k_0}. \tag{33}$$

Calculations of the ratio $\overline{x_+}/\overline{x_{\mathrm{R}}}$ are presented in Table I, showing that $\overline{x_+}$ overestimates $\overline{x_{\mathrm{R}}}$ by only 1%–4% for the materials considered, which is more accurate than any of the other estimates considered above. Values of β_+ are listed in the last column (which correspondingly underestimate the magnitude of β_x by 1%–4%). Pyrex is unique in this group in that its nonlinearity coefficient is negative, indicating that the horizontal velocity component steepens in a direction opposite that for the other materials (and therefore the direction of the impulse formed in the vertical velocity component is reversed in sign). The nonlinearity coefficient for Rayleigh waves varies from order unity for homogeneous materials to order 10^3 for rocks containing cracks, granularity, and other microinhomogeneous features, as is the case for the nonlinearity coefficient associated with compression waves.[16] In particular, Johnson and Rasolofosaon[14] calculated β_l for compression wave propagation in the last seven materials listed in Table I (use the relation $\beta_l=-\beta$ to convert between their notation and ours). Each of their values for β_l is approximately two to three times our corresponding value for β_+.

IV. CONCLUSION

Whereas the coefficient of nonlinearity for the propagation of finite-amplitude sound is a constant, it was shown that a consistently defined nonlinearity coefficient for Rayleigh waves is strongly frequency dependent and weighted toward sum rather than difference frequency generation. It is specifically this frequency dependence that distinguishes the combination of local and nonlocal nonlinearity in Rayleigh waves from the exclusively local nonlinearity in sound waves. Numerical calculations revealed that local nonlinearity dominates nonlocal nonlinearity in the distortion process leading up to shock formation. The above observations were used to develop analytical expressions for the shock formation distance, Eq. (33), and effective constant coefficient of nonlinearity, Eq. (32), that, to within a few percent, approximate numerical calculations of the shock formation distance for nonlinear Rayleigh wave propagation in a wide variety of materials.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation, the Office of Naval Research, the Schlumberger Foundation, and the Department of Energy through the Earth and Environmental Sciences Division at Los Alamos National Laboratory.

APPENDIX A: NUMERICAL CALCULATIONS

The calculations of $\overline{x_{\mathrm{R}}}$ reported in Sec. III were obtained by solving the following discrete form of Eq. (1) for the spectral components $\tilde{v}_n(x)$ at the harmonically related frequencies $n\omega_0$:[2,7]

$$\frac{d\tilde{v}_n}{dx}+\alpha_n\tilde{v}_n=\frac{\mu k_0 n^2}{2\rho c_{\mathrm{R}}^3\zeta}\left(2\sum_{m=n+1}^{\infty} R_{m,n-m}\tilde{v}_m\tilde{v}^*_{m-n} - \sum_{m=1}^{n-1} R_{m,n-m}\tilde{v}_m\tilde{v}_{n-m}\right), \tag{A1}$$

where asterisks designate complex conjugates ($\tilde{v}_{-n}=\tilde{v}_n^*$), and losses were introduced *ad hoc*, with α_n the small-signal attenuation coefficient at frequency $n\omega_0$. As discussed below, a small amount of attenuation is needed for numerical stability when shocks are formed. We are interested here in

calculating the horizontal velocity waveform $v_x(x,\tau)$ at the interface, which is given by

$$v_x(x,\tau)=(\xi_t+\eta)\,\frac{i}{2}\sum_{n=-\infty}^{\infty}\frac{n}{|n|}\,\tilde{v}_n(x)e^{-in\omega_0\tau}. \tag{A2}$$

To render Eq. (A1) dimensionless for numerical integration we introduce the following variables:

$$V_n=\tilde{v}_n/v_0,\quad X=x/x_0,\quad A_n=\alpha_n x_0, \tag{A3}$$

where $v_0=v_{x0}/|\xi_t+\eta|$, v_{x0} is the horizontal velocity amplitude in Eq. (22), and x_0 is defined in Eq. (27). Substitution of Eqs. (A3) into Eq. (A1) yields

$$\frac{dV_n}{dX}+A_nV_n=2n^2\sum_{m=n+1}^{N}R_{m,n-m}V_mV_{m-n}^{*}-n^2\sum_{m=1}^{n-1}R_{m,n-m}V_mV_{n-m}. \tag{A4}$$

Only a finite number of harmonics, N, is retained for computational reasons. We also express the horizontal velocity waveform in dimensionless form:

$$V_x=v_x/v_{x0},\quad \theta=\omega_0\tau, \tag{A5}$$

in terms of which Eq. (A2) becomes

$$V_x(X,\theta)=\frac{1}{2i}\sum_{n=-N}^{N}\frac{n}{|n|}\,V_n(X)e^{-in\theta}. \tag{A6}$$

The boundary condition in Eq. (22) thus corresponds to $V_1=-1$, and $V_n=0$ for $n>1$, at $X=0$.

To integrate the system of coupled ordinary differential equations (A4) a standard fourth-order Runge–Kutta routine with adjustable step-size control was used. A quadratic frequency dependence for the attenuation coefficient was assumed in all numerical calculations ($A_n=n^2A_1$).[17]

Shock formation is said to occur when a discontinuity (infinite derivative) first appears in the time waveform. However, thermoviscous attenuation in real materials prevents the waveform from becoming discontinuous—the shocks have finite rise times, and the rise times increase with attenuation. What we are seeking is the shock formation distance in an ideal, lossless solid. The reason for including attenuation in the calculations is that it is required for numerical stability in the shock wave region.[1,2] We therefore wish to keep attenuation sufficiently small (i.e., small A_1) that a reasonable approximation of lossless propagation is obtained up to shock formation, yet not so small that an unreasonable number of harmonics must be retained in the calculations.

For most of the materials considered in Table I (i.e., all but Pyrex), the nonlinearity corresponds in sign to that which is normally encountered in fluids. That is, positive portions of the horizontal velocity waveform propagate faster than the small signal speed c_R, negative portions propagate slower. For the source condition in Eq. (22), these waveforms develop shocks precisely at $\tau=0$. For opposite nonlinearity with the same source condition, the shocks develop at $\tau=\pi/\omega_0$. Proximity to shock formation may be measured in terms of the slope of the waveform at $\tau=\tau_s$ (i.e., $\theta=\theta_s$, $\theta_s=\omega_0\tau_s$), where τ_s designates the shock location. We thus introduce the function

$$\Theta(X)=1\Big/\left|\frac{\partial V_x}{\partial\theta}\right|_{\theta=\theta_s}, \tag{A7}$$

normalized such that $\Theta(0)=1$ for our sinusoidal source condition. The shock formation distance $\overline{x_R}$ is defined as the location where an infinite slope first appears in the waveform, which corresponds to the point where $\Theta=0$.

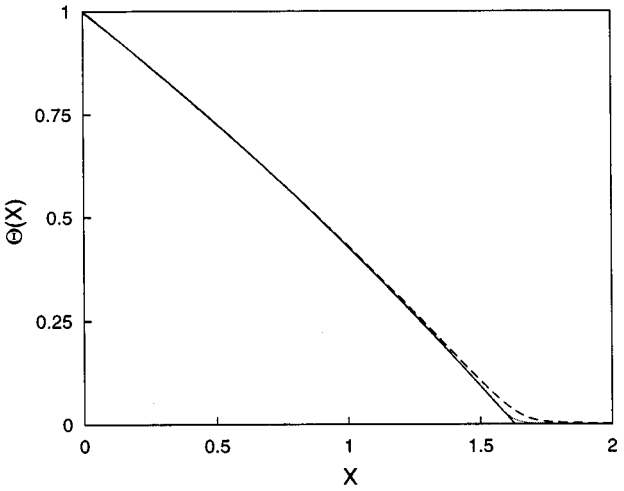


FIG. A1. Numerical calculation of the reciprocal of the maximum time derivative on the horizontal velocity waveform as a function of distance, with $A_1=10^{-3}$ and $N=300$ (dashed line, ---) and with $A_1=10^{-4}$ and $N=500$ (dotted line, ···). The solid line (—) is a polynomial approximation of the second case.

To demonstrate this procedure, we calculate the shock formation distance for steel using the same parameters as in Fig. 1. In Fig. A1, $\Theta(X)$ is plotted for $A_1=10^{-3}$ (dashed line, $N=300$) and $A_1=10^{-4}$ (dotted line, $N=500$). The two curves indicate that shock formation occurs somewhere in the region $1.6<X<1.7$. The finite shock thickness, due to both thermoviscous attenuation and numerical error, results in the nonzero value of Θ.

To obtain a more precise estimate of the shock formation distance, the following method was used. The ''plateau'' part of the curve for Θ was discarded (this is where absorption plays a very important role, e.g., $X\gtrsim 1.6$ for the case with $A_1=10^{-4}$), and the remainder of the curve was fitted with a polynomial. It was found that good agreement is achieved with a second-order polynomial. The fitting curve for the case with $A_1=10^{-4}$ was found to be $\Theta=-0.067X^2-0.503X+0.997$, which is plotted as the solid line in Fig. A1. The intersection of the fitting curve with the X axis is our numerical estimate of the shock formation distance $\overline{x_R}$, and we thus obtain $\overline{x_R}=1.63x_0$ for steel.

APPENDIX B: CALCULATION OF $\bar{f}_+$

We begin by repeating Eq. (31):

$$\overline{f_+}=\frac{2}{\pi}\int_0^{\pi/2}f_\phi(\phi)\,d\phi, \tag{B1}$$

where $f_\phi(\phi)$ is an alternative (polar) form of $F(\omega_1,\omega_2)$, defined in Eq. (14). In the region of interest, $0\leqslant\phi\leqslant\pi/2$, $f_\phi(\phi)$ can be written as

$$f_\phi(\phi)=\alpha'\left(\frac{1}{\xi_t+\xi_l}+\frac{1+\tan\phi}{2\xi_t+(\xi_t+\xi_l)\tan\phi}+\frac{1+\tan\phi}{\xi_t+\xi_l+2\xi_t\tan\phi}\right)+\beta'\left(\frac{1}{\xi_t+\xi_l}+\frac{1+\tan\phi}{2\xi_l+(\xi_t+\xi_l)\tan\phi}+\frac{1+\tan\phi}{\xi_t+\xi_l+2\xi_l\tan\phi}\right)+\frac{3\gamma'}{2\xi_l}, \tag{B2}$$

where Eq. (14) was used. Substitution of Eq. (B2) into Eq. (B1) and integration yield

$$\overline{f_+}=\alpha'\left(\frac{1}{\xi_t+\xi_l}+2\frac{3\xi_t+\xi_l}{4\xi_t^2+(\xi_t+\xi_l)^2}+\frac{4}{\pi}\frac{\xi_l-\xi_t}{[4\xi_t^2+(\xi_t+\xi_l)^2]}\ln\frac{(\xi_t+\xi_l)}{2\xi_t}\right)+\beta'\left(\frac{1}{\xi_t+\xi_l}+2\frac{3\xi_l+\xi_t}{4\xi_l^2+(\xi_t+\xi_l)^2}+\frac{4}{\pi}\frac{\xi_t-\xi_l}{[4\xi_l^2+(\xi_t+\xi_l)^2]}\ln\frac{(\xi_t+\xi_l)}{2\xi_l}\right)+\frac{3\gamma'}{2\xi_l}. \tag{B3}$$

The following integrals were used to calculate Eq. (B3):

$$\int\frac{dx}{a+b\tan cx}=\frac{1}{c(a^2+b^2)}[acx+b\ln(a\cos x+b\sin cx)],$$

$$\int\frac{dx}{a+b\cot cx}=\frac{1}{c(a^2+b^2)}[acx-b\ln(a\sin x+b\cos cx)].$$

[1] E. A. Zabolotskaya, ''Nonlinear propagation of plane and circular Rayleigh waves in isotropic solids,'' J. Acoust. Soc. Am. **91**, 2569–2575 (1992).

[2] D. J. Shull, M. F. Hamilton, Yu. A. Il'insky, and E. A. Zabolotskaya, ''Harmonic generation in plane and cylindrical nonlinear Rayleigh waves,'' J. Acoust. Soc. Am. **94**, 418–427 (1993).

[3] D. T. Blackstock, ''Propagation of plane sound waves of finite amplitude in nondissipative fluids,'' J. Acoust. Soc. Am. **34**, 9–30 (1962).

[4] O. V. Rudenko and S. I. Soluyan, *Theoretical Foundations of Nonlinear Acoustics* (Plenum, New York, 1977).

[5] M. F. Hamilton, Yu. A. Il'insky, and E. A. Zabolotskaya, ''Local and nonlocal nonlinearity in Rayleigh waves,'' J. Acoust. Soc. Am. **97**, 882–890 (1995).

[6] M. F. Hamilton, Yu. A. Il'insky, and E. A. Zabolotskaya, ''Evolution equations for nonlinear Rayleigh waves,'' J. Acoust. Soc. Am. **97**, 891–897 (1995).

[7] E. Yu. Knight, M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, ''General theory for the spectral evolution of nonlinear Rayleigh waves,'' J. Acoust. Soc. Am. **102**, 1402–1417 (1997).

[8] D. T. Blackstock, ''Nonlinear Acoustics (Theoretical),'' in *American Institute of Physics Handbook* (McGraw-Hill, New York, 1972), 3rd ed., Chap. 3n.

[9] R. T. Beyer, ''Parameter of nonlinearity in fluids,'' J. Acoust. Soc. Am. **32**, 719–721 (1960).

[10] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity* (Pergamon, New York, 1986), 3rd ed., pp. 106–107.

[11] F. D. Murnaghan, *Finite Deformation of an Elastic Solid* (Chapman and Hall, New York, 1951), pp. 63–64.

[12] F. H. Fenlon, ''A recursive procedure for computing the nonlinear spectral interactions of progressive finite-amplitude waves in nondispersive fluids,'' J. Acoust. Soc. Am. **50**, 1299–1312 (1971).

[13] R. F. S. Hearmon, ''The elastic constants of crystals and other anisotropic materials; The third- and higher-order elastic constants,'' in *Landolt–Börnstein: Numerical Data and Functional Relationships in Science and Technology*, edited by K.-H. Hellwege and A. M. Hellwege (Springer, New York, 1979), Group III, Vol. 11, pp. 245–286.

[14] P. A. Johnson and P. N. J. Rasolofosaon, ''Nonlinear elasticity and stress-induced anisotropy in rocks,'' J. Geophys. Res. **101**, 3113–3124 (1996). Note that their coefficient of nonlinearity β equals the negative of our Eq. (9) for β_l.

[15] E. Yu. Knight, ''Generalization of the theory for nonlinear Rayleigh waves to nonplanar and transient waveforms, and investigation of pulse propagation,'' M.A. thesis, The University of Texas at Austin, 1995.

[16] L. A. Ostrovsky, ''Wave processes in media with strong acoustic nonlinearity,'' J. Acoust. Soc. Am. **90**, 3332–3337 (1991).

[17] R. W. Lardner, ''Nonlinear Rayleigh waves: Harmonic generation, parametric amplification, and thermoviscous damping,'' J. Appl. Phys. **55**, 3251–3260 (1984).

Generation of an enhanced acoustic streaming in a bounded space

Yorinobu Murata
Department of Opto-Mechatronics, Faculty of Systems Engineering, Wakayama University, 930 Sakaedani, Wakayama 640, Japan

Satoshi Kimura, Seiji Takei, Keiji Minagawa, Jun-ichi Takimoto, and Kiyohito Koyama
Department of Materials Science and Engineering, Faculty of Engineering, Yamagata University, 4-3-16 Jonan, Yonezawa 992, Japan

(Received 20 November 1994; accepted for publication 22 July 1997)

A procedure to generate enhanced acoustic streaming into a bounded space has been studied. Under continuous ultrasound radiation, it was difficult to generate enhanced acoustic streaming in bounded space because of disturbances by reflected waves. Efforts were made to remove the disturbances by adjusting the repeat time of tone bursts of ultrasound. An experimental system has been developed for visualizing acoustic streaming. As a consequence of experiments, the streaming velocity in the bounded space has been enhanced by about 2.5 times by using a series of tone bursts of ultrasound.
 [S0001-4966(97)01411-2]

PACS numbers: 43.25.Nm [MAB]

When intense ultrasound propagates in fluid, streaming is generated along the acoustic axis in the medium. This phenomenon is called acoustic streaming. The generation mechanism of acoustic streaming has been explained as a radiation pressure gradient due to the absorption of the acoustic wave in the propagation process.[1] The velocity of the acoustic streaming generally depends on the radiation pressure gradient. The velocity becomes greater at higher frequency because absorption increases. Examples of acoustic streaming in some unbounded spaces filled by media, such as air[2] and water,[3] have been reported. However, efficient generation and practical applications of acoustic streaming have never been reported.

There are various possible applications of acoustic streaming. The most promising application is micro-mixing of materials. Detailed regulation of the flow behavior should be useful for obtaining materials with fine structures and new functions. The advantage of utilizing acoustic streaming is that a specific part of the material can be selectively stirred without any mechanical stirrer. This allows us to control the mixing of the samples from outside of the cell. We have been trying to apply acoustic mixing to polymer processing in order to improve the quality of molded objects. If polymeric materials in the mold are mixed, defective molded objects can be decreased because the degree of orientation in the materials is lowered. When the acoustic streaming is generated from outside the mold at high temperature, materials in the mold can be mixed efficiently. To generate acoustic streaming in the mold, the propagation of ultrasound in a bounded space must be considered. When a continuous wave is radiated into a bounded space, the radiated wave interferes with waves reflected from the bounding surfaces of the space. Therefore, a radiation pressure gradient is difficult to create in a medium sandwiched between bounding surfaces.

In this paper, we report a procedure to generate enhanced acoustic streaming in a bounded space. A new experimental system has been developed which enables the visualization of the acoustic streaming in a bounded space. Streaming velocities in the medium under various conditions of radiating ultrasound are discussed on the basis of visualization experiments.

The visualization experiments in bounded space were carried out using the system shown in Fig. 1. A silicone oil (KF50 Shin-Etsu Chemical Co., Ltd.) was used as the medium. The silicone oil has characteristics suitable for this experiment, i.e., good transparency and easy generation of the acoustic streaming. The viscosity, the sound absorption coefficient, and the sound velocity of the silicone oil were, respectively, 100 cpoise, 0.6×10^{-14} neper$\cdot$s$^2\cdot$cm^{-1}, and 980 m/s. A piezoelectric ceramic (PZT) transducer was used. The diameter and the resonance frequency of the transducer were 15 mm and 1.1 MHz, respectively. Continuous or tone bursts of a sinusoidal signal were generated by a function generator (Hewlett-Packard 8116A) to drive the transducer. The frequency and the voltage were 1.1 MHz and 165 $V_{p\text{-}p}$. The number of waves in the tone bursts of ultrasound was constant at 40, which almost corresponded to the height of the medium when the number was calculated in wavelengths. Generated ultrasound was radiated into the medium through a buffer and a metal plate made of iron having a sound velocity of about 5200 m/s. The thickness of the metal plate was 15 mm. The buffer was made of quartz glass having a sound velocity of about 5400 m/s, and was used for protecting the transducer from heat when a high temperature medium such as a melted polymer was used. The length and the diameter of the buffer were 63 and 26 mm, respectively. The cell was a transparent cylinder made of glass. The height and the diameter of the cell were 150 and 70 mm, respectively. The parts (the transducer, the buffer, the plate, and the cell) were set together by using epoxy resin adhesive. The adhesive did not affect ultrasound propagation because the adhesive layer was very thin. The silicone oil was filled to 20 mm above the bottom of the cell. To visualize the acoustic streaming, we dispersed high-density polyethylene (HDPE)

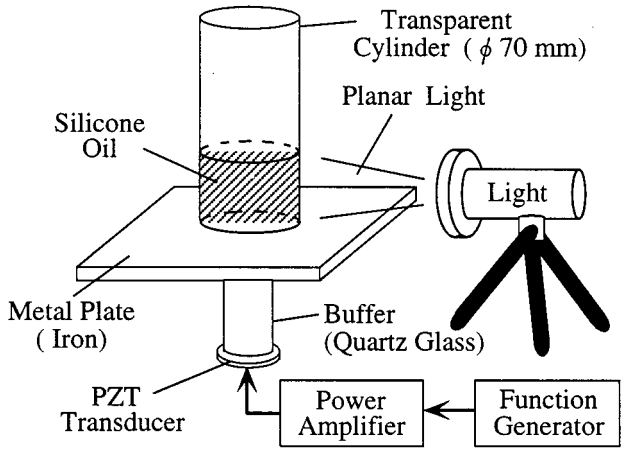


FIG. 1. Experimental system for visualizing acoustic streaming in a bounded space.

powder, having a diameter of about 300 μm, in the medium and had a planar light illuminating the axial plane of the transducer. The experiments were carried out at room temperature. After the acoustic streaming had reached a stable state, the streaming was visualized by photographing it for an exposure time of 10 s. The streaming velocities along the acoustic axis were measured by recording the movement of the particles using a video camera.

Figure 2(a) represents the visualized photograph under no radiation; Fig. 2(b) under continuous ultrasound radiation; and Fig. 2(c) with tone bursts of ultrasound. In Fig. 2(c), the repeat time of the tone burst signal was 110 μs. In the photographs of Fig. 2, the cloudy parts appearing at the bottom of the cell were clumps of the HDPE powder. Photographing through a curved surface, the side of the cell distorted the picture. The clumps actually collected at the corner of the bottom in the cell in the experiment. Sketches in Fig. 2 do not represent the streaming pattern traced on the photograph but represent that of the hatched area in Fig. 1. The arrows in the sketches show the partial direction of the streaming. The length of the lines do not represent the actual streaming velocity. It was found in Fig. 2(b) that only a weak acoustic streaming was generated by the continuous radiation. On the other hand, the acoustic streaming under the tone burst streaming was more obvious [see Fig. 2(c)]. Moreover, the directivity of the streaming under the tone bursts of ultrasound was sharper. The streaming velocity along the acoustic axis under the tone bursts of ultrasound was faster (maximum value 1 mm/s) than that under the continuous ultrasound (maximum value 0.4 mm/s). The reason for the weak streaming under continuous radiation is schematically explained in Fig. 3. For continuous ultrasound [see Fig. 3(a)], the radiation pressure gradient in the medium is disturbed by the interference between the transmitted wave and the wave reflected by the top surface, the boundary surface between the medium and air. This was evident because continuous ultrasound generated a strong streaming in the cell when the interference was reduced by filling the cell with liquid. Interference in the medium can be removed by using an appropriate tone burst length [see Fig. 3(b)]. Enhanced acoustic streaming can be generated when we transmit tone bursts having a repeat time which nearly equals the time required for the reflected wave to fade away. The repeat time then depends on the height and the absorption coefficient of the medium.

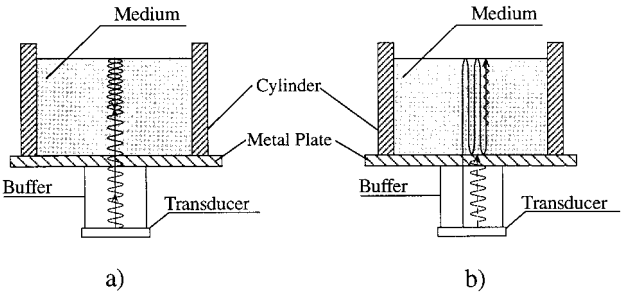


FIG. 3. Schematic representations of ultrasound propagation in bounded space (a) under continuous ultrasound radiation and (b) under tone bursts. In (b), the solid line represents the propagation of the tone bursts transmitted at the first, and the wavy line represents that transmitted after 110 μs corresponding to the best repeat time.

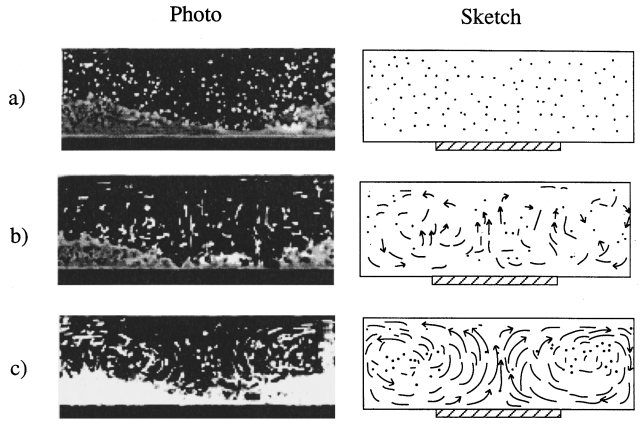


FIG. 2. Visualization of the acoustic streaming (a) under no radiation, (b) under continuous ultrasound, and (c) under tone bursts. The number of waves and the repeat time of the tone bursts signal were 40 and 110 μs, respectively. The right sketches show the streaming patterns of the hatched area. The arrows in the sketches show the partial direction of the streaming. The hatched part represents the position of the buffer.

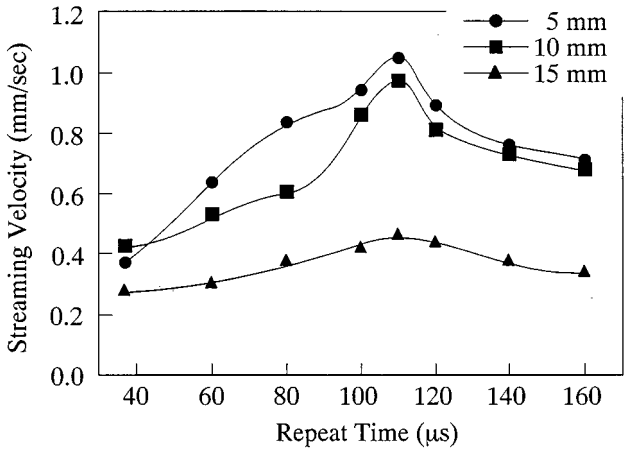


FIG. 4. Variation of the streaming velocity along the acoustic axis with the repeat time of the tone bursts. The parameter from one curve to the next is the distance from the bottom of the cell.

Figure 4 shows variation of the streaming velocity with repeat time of the tone bursts for different container heights (frequency). The number of waves in the tone burst was a constant 40. Tone bursts having a repeat time of less than 40 μs were essentially equal to continuous ultrasound. The streaming velocity along the acoustic axis showed a maximum value when the repeat time was 110 μs. This repeat time approximately corresponded to the sum of 2.33 round trips in the medium and the time for a single pass of sound from the transducer to the cell bottom. It was considered that the reflected wave had attenuated during the propagation of about 2.33 round trips in the medium. The reason for having a peak velocity at 110 μs in Fig. 4 was considered as follows. When the repeat time is too short, the reflected waves affect the acoustic streaming like continuous ultrasound. When the repeat time is long, on the other hand, the average acoustic energy per period decreases. Then the radiation pressure is smaller. Therefore, to generate an enhanced acoustic streaming in a bounded space, the repeat time of the tone burst must be adjusted according to the thickness and the attenuation of the medium. The number of waves in tone bursts was a constant throughout the experiments, but should be adjusted according to the thickness of the sample in the other applications. It was learned by experiment that not only the velocity but also the directivity of acoustic streaming can be controlled by adjusting the repeat time of the tone bursts. The proposed procedure therefore partially mixes the medium in a cell.

In conclusion, we have tried to enhance acoustic streaming in a bounded space. A system for visualizing the acoustic streaming was developed to measure the streaming velocity under various conditions. Enhanced acoustic streaming in a cell has been achieved by using a series of tone bursts instead of continuous ultrasound. Acoustic streaming in various shapes of cells can be controlled by adjusting the repeat time of the tone bursts.

ACKNOWLEDGMENT

This work was partly supported by the Grant-in-Aid from the Ministry of Education, Science and Culture, Japan.

[1] V. A. Shutilov, *Fundamental Physics of Ultrasound* (Gordon and Breach Science, New York, 1988).
[2] H. Mitome, ''Ultrasonic Levitation and Accompanying Acoustic Streaming,'' Jpn. J. Appl. Phys., Part 1 **28**, 146–148 (1989).
[3] H. Mitome, ''Study of the Generation Mechanism of an Acoustic Jet Through Visualization Experiments,'' Jpn. J. Appl. Phys., Part 1 **30**, 60–62 (1991).

Nonlinear distortion of short pulses radiated by plane and focused circular pistons

Michalakis A. Averkiou
Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

Mark F. Hamilton
Department of Mechanical Engineering, The University of Texas at Austin, Austin, Texas 78712-1063

(Received 19 June 1996; accepted for publication 2 June 1997)

Detailed measurements of finite-amplitude pulses radiated by plane and focused circular pistons in water are presented. Comparisons of time waveforms and frequency spectra, both on and off axis, are made with numerical calculations based on the nonlinear parabolic wave equation. Emphasis is on nonlinear distortion of amplitude- and frequency-modulated tone bursts. Use of short pulses enabled resolution of the direct and diffracted waves prior to their coalescence and subsequent shock formation along the axis of the source. Because of its relevance to investigations of cavitation inception, attention is devoted to variation of the peak positive (p_+) and negative (p_-) pressures along the axis of a focused source. It is shown that with increasing source amplitude, the maximum of each shifts away from the focal plane, toward the source. This effect is more pronounced for p_- than for p_+. © *1997 Acoustical Society of America.* [S0001-4966(97)06110-9]

PACS numbers: 43.25.Ts, 43.80.Sh [MAB]

INTRODUCTION

The use of intense ultrasound in medical and industrial applications has increased considerably in recent years. Both plane and focused sources are used widely in either continuous wave (cw) or pulsed mode, and at intensities which lead to nonlinear effects such as harmonic generation and shock formation. Typical ultrasonic sources generate strong diffraction phenomena, which combine with finite amplitude effects to produce waveforms that vary from point to point within the sound beam. Nonlinear effects have become especially important at acoustic intensities employed in many current therapeutic and surgical procedures.[1] In addition, biological media can introduce significant absorption of sound, which must also be considered.

The emphasis of the present article is on detailed comparisons of theory and experiment for pulsed finite-amplitude sound beams radiated from plane and focused circular pistons. Comparisons of theory and experiment for the case of small-signal transient radiation from plane circular pistons have been reviewed by Hutchins and Hayward.[2] Similar comparisons for the case of focused circular pistons have also been reported, with perhaps the first extensive investigation of this kind described in the article by Goodsitt *et al.*[3] More recently, Djelouah *et al.*[4,5] compared theory with measurements of sufficiently short pulses that the direct and diffracted wave contributions along the axis were easily resolved.

The first comparisons of an appropriate theoretical model with measurements of pulsed finite-amplitude sound beams containing shocks are presented in the work by Baker and Humphrey.[6–8] They compared measurements of short pulses (one cycle excitation that resulted in roughly three cycles due to the bandwidth of the source) at finite amplitudes with results from a finite difference algorithm that solves the KZK (Khokhlov–Zabolotskaya–Kuznetsov) nonlinear parabolic wave equation.[9,10] They investigated pulses radiated by plane circular pistons in Ref. 6, by focused circular pistons in Ref. 7, and by both varieties in Ref. 8. A frequency domain algorithm was used to solve the KZK equation, which required repetition of each pulse periodically in time to represent the signal in terms of a discrete Fourier series.

The purpose of the present article is to compare measurements of pulsed finite amplitude sound beams in water, both focused and unfocused, with predictions obtained from a time domain algorithm for solving the KZK equation. The investigation also differs from those reported by Baker and Humphrey[6–8] principally in terms of greater breadth and precision. Both amplitude- and frequency-modulated tone bursts are considered, coalescence and interference of direct and edge waves are reported, and the likelihood of preferential sites for acoustic cavitation[11] in a focused sound beam is investigated with measurements and calculations of negative pressure excursions along the axis of the beam.

I. THEORY

The theoretical predictions in Sec. III are obtained from numerical solutions of the KZK nonlinear parabolic wave equation, written here in terms of the sound pressure p:

$$\frac{\partial^2 p}{\partial z\,\partial t'} = \frac{c_0}{2}\left(\frac{\partial^2 p}{\partial r^2} + \frac{1}{r}\frac{\partial p}{\partial r}\right) + \frac{\delta}{2c_0^3}\frac{\partial^3 p}{\partial t'^3} + \frac{\beta}{2\rho_0 c_0^3}\frac{\partial^2 p^2}{\partial t'^2}. \tag{1}$$

Only axisymmetric radiation is considered, with z the coordinate along the axis of the sound beam, r the distance from the beam axis, and $t' = t - z/c_0$ a retarded time, where c_0 is the propagation speed. The first term on the right-hand side of Eq. (1) accounts for diffraction. Thermoviscous attenuation is taken into account by the second term, where δ is the sound diffusivity.[12] In the third term, $\beta = 1 + B/2A$ is the

coefficient of nonlinearity[13] and ρ_0 is the ambient density of the fluid. In general, Eq. (1) is an accurate model of the sound field produced by directional sound sources ($ka \gg 1$, where k characterizes the wave number and a the radius of the source) at distances beyond a few source radii and in regions close to the axis of the source, the paraxial region (up to about 20° off the z axis in the far field). These restrictions are satisfied in most practical applications of directional sound beams. A complete discussion of the domain of validity of the KZK equation for plane and focused piston sources is provided in Refs. 14 and 15.

Two source conditions are considered, one for plane and the other for focused circular radiators, each with uniform velocity amplitude and radius a. For the plane circular sources we prescribe the boundary condition

$$p = p_0 E(t) \sin[\omega_0 t + \phi(t)] H(1 - r/a) \quad \text{at } z = 0, \tag{2}$$

where p_0 is the effective pressure amplitude at the source, $E(t)$ and $\phi(t)$ are amplitude and phase modulations, respectively, of a signal at angular frequency ω_0 (in all calculations and experiments, $f_0 = \omega_0/2\pi$ is the resonance frequency of the transducer), and H is the Heaviside unit step function. Since the order of accuracy of the KZK equation is consistent with the use of the linear plane wave impedance relation[16] $p = \rho_0 c_0 u$, where u is the particle velocity in the z direction, we have used this relation to express the boundary condition (2) in terms of pressure rather than particle velocity. For the envelope function we define

$$E(t) = \exp[-(2t/T)^{2m}], \tag{3}$$

where T characterizes its duration and m its rise and decay time. The instantaneous angular frequency of the signal is $\Omega(t) = \omega_0 + d\phi/dt$.

Numerical solutions of Eq. (1) subject to Eq. (2) were obtained using finite difference algorithms. With the exception of the predictions in Fig. 2 for continuous radiation from a plane circular piston, which were calculated with the spectral code described by Naze Tjøtta *et al.*,[15] all calculations were performed with the time domain code described by Lee and Hamilton.[17] An advantage of the spectral code is that propagation in media with arbitrary attenuation and dispersion can be modeled. However, only thermoviscous attenuation, as described by Eq. (1), need be considered for the experiments reported in the present article. As discussed by Cleveland *et al.*,[18] absorption and dispersion due to multiple relaxation phenomena can also be included in the time domain code. An advantage of the time domain code is its suitability for modeling pulses, random signals, and shock rise times. An early listing of the time domain code is provided by Lee.[19]

Given the boundary condition in Eq. (2), the solutions of Eq. (1) are determined completely (numerical error notwithstanding) by the following two independent dimensionless parameters:

$$A_u = \alpha_0 z_0, \quad N_u = z_0/\overline{z}. \tag{4}$$

The subscript u, and subsequently f, is used to distinguish between the parameters for unfocused (plane) and focused sources, respectively. Both parameters in Eqs. (4) are normalized by the Rayleigh distance $z_0 = \omega_0 a^2/2c_0$. In the definition of the absorption parameter A_u, $\alpha_0 = \delta\omega_0^2/2c_0^3$ is the thermoviscous attenuation coefficient at frequency ω_0. Thus, A_u is the ratio of the characteristic diffraction and absorption length scales. In the nonlinearity parameter N_u, $\overline{z} = \rho_0 c_0^3/\beta\omega_0 p_0$ is the plane wave shock formation distance, and therefore N_u is the ratio of the characteristic diffraction and nonlinearity length scales.

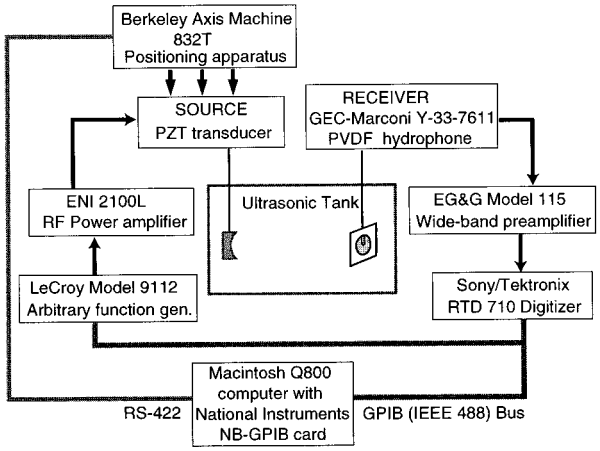


FIG. 1. Block diagram of the experimental setup.

For a focused source with focal length d we replace t with $t + r^2/2c_0 d$ in Eq. (2). The code as listed by Lee[19] for plane, diverging sound beams was modified to accommodate the geometry of focused beams. Three dimensionless parameters must now be specified:

$$G = z_0/d, \quad A_f = \alpha_0 d, \quad N_f = d/\overline{z}, \tag{5}$$

where G is the linear focusing gain, equal to the peak value of p/p_0 at the geometric focus $(r,z) = (0,d)$ for continuous small-signal radiation at frequency ω_0 in a lossless fluid. The absorption parameter A_f and nonlinearity parameter N_f are scaled here according to the focal length.

II. EXPERIMENT

A block diagram of the experimental setup is shown in Fig. 1. The measurements were made in the ultrasonics water tank described previously by TenCate[20] (see also Ref. 21). Positioning and movement of the sources with respect to the hydrophone was accomplished with a Berkeley Axis Machine 832T (20-μm spatial resolution).

Two different plane sources were used, both made by Panametrics. One has an effective radius $a = 9.4$ mm and a center resonance frequency $f_0 = 2.25$ MHz, and the other has an effective radius $a = 12.1$ mm and a center resonance frequency $f_0 = 1$ MHz. The effective radius of the source was determined by comparing measurements of far-field beam patterns with linear theory. The corresponding Rayleigh distances are $z_0 = 420$ mm and $z_0 = 310$ mm, respectively (the measured sound speed in the water was $c_0 = 1486$ m/s).

We used the same focused source, manufactured by Harisonic, as in experiments on cw radiation reported earlier.[22] Focusing was achieved not with a lens, but by hav-

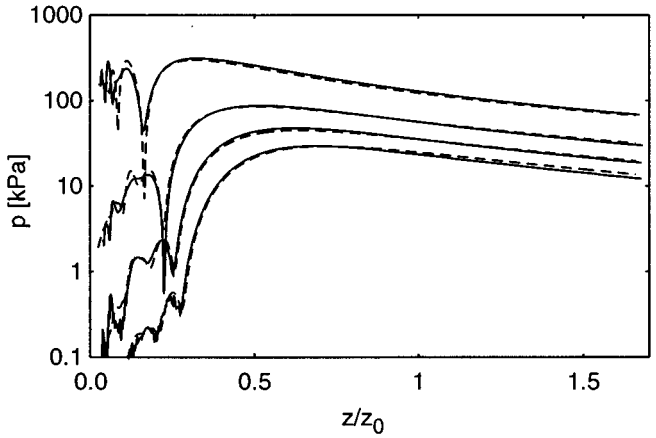


FIG. 2. Comparison of experiment (solid lines; $f_0=2.25$ MHz, $a=9.4$ mm, $z_0=420$ mm) and theory (dashed lines; $A_u=0.053$, $N_u=0.97$) for the pressure amplitudes of the lowest four harmonic components (fundamental through fourth harmonic in top to bottom order) in a plane piston beam.

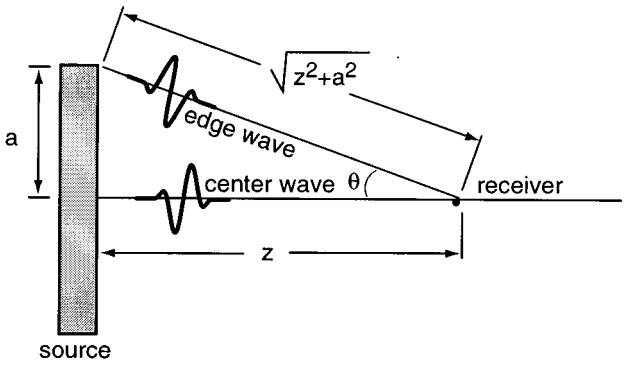


FIG. 3. Geometry of direct and diffracted wave arrivals.

ing the PZT transducer cut to form a spherical concave surface. As described in the previous work,[22] the effective radius and focal length of the source were determined to be $a=18.8$ mm and $d=160$ mm, respectively. When driven continuously at 2.25 MHz, the focusing gain is $G=10.5$.

A LeCroy model 9112 arbitrary function generator was used to generate the source waveforms, and an ENI 2100L RF power amplifier was used to amplify the signal to the sources. The receive system is identical to that employed in the earlier experiments.[22] We used a bilaminar Marconi PVDF membrane hydrophone (1-mm active diameter, 25-μm film thickness). This type of hydrophone has a resonance frequency at about 23 MHz, and its sensitivity falls off above that frequency.[23] The hydrophone was calibrated at National Physical Laboratory (Teddington, UK), and the reported variation of sensitivity with frequency was factored out of the measurements. The hydrophone signal was amplified with an EG&G model 15 wide-band preamplifier (50-MHz cutoff frequency), and sampled with a Sony/Tektronix RTD 710 digitizer (10 bit) at a rate of 200 MHz. Dynamic range was improved by repeated averaging of the received signals, typically 1024 times.

III. RESULTS

In this section we present our experimental results and compare them with theoretical predictions. We divide our results in two categories, plane pistons and focused pistons. We begin with the plane pistons.

A. Plane pistons

In order to compare measured waveforms with the theoretical model we first ensure that the source behaves as a circular piston, as assumed in accordance with Eq. (2). In Fig. 2 we show measurements of an axial propagation curve produced by continuous radiation from the plane 2.25-MHz source (solid lines), and theoretical predictions (dashed lines) based on a code that solves the KZK equation in the frequency domain[15] (all other calculations in this article are obtained from a time domain code[17]). The absorption and nonlinearity parameters used in the theoretical model are $A_u=0.053$ and $N_u=0.97$, respectively. The first four harmonics are shown. From Fig. 2 we see that there is excellent agreement between measurement and theory over a dynamic range of about 70 dB. Measured axial propagation curves for the 1-MHz source exhibited similar agreement with theory. We thus conclude that our sources behave as nearly ideal pistons, and we proceed to our measurements of pulses.

A circular piston is excited with a short pulse (a couple of cycles), and the field at a distance z along the axis of the source is monitored as shown in Fig. 3. Very near the source, two distinct contributions to the received signal can be identified. The first is called the direct wave (or center wave), because its arrival time is equal to that of a signal originating from the exact center of the source. The second is called the diffracted wave (or edge wave), because its arrival time corresponds to that of a wave radiated at the edge of the source. The direct wave arrives first after traveling a distance z, and the diffracted wave, an inverted replica of the direct wave, arrives second after traveling a distance $\sqrt{z^2+a^2}$. The difference in arrival times approaches zero far away from the source. Small-signal theory and experiments related to these effects are reviewed by Hutchins and Hayward.[2]

In Fig. 4 we show measured time waveforms for a short pulse at various positions along the axis of the source. The 2.25-MHz source was excited with a single cycle at that frequency, but due to bandwidth limitations of the source the waveform actually radiated was longer (approximately 2 cycles). At $z=2$ mm, the measurement location nearest to the source, the direct wave is clearly seen (first arrival), but the edge wave (second arrival) is difficult to distinguish. In general, as observed by others,[2] the measured edge waves shown in the first column change considerably from one location to the next because of the directivity of the hydrophone. The angles $\theta=\arctan(a/z)$ given in the figure correspond to the angle formed by the axis of the source and the line that passes through the axial observation point and the edge of the circular source (see Fig. 3). As indicated by the measurements of directivities reported by Bacon,[23] the sensitivity of a membrane hydrophone having an active element 1 mm in diameter may be expected to vary dramatically at 2.25 MHz through the range of angles in the first column of

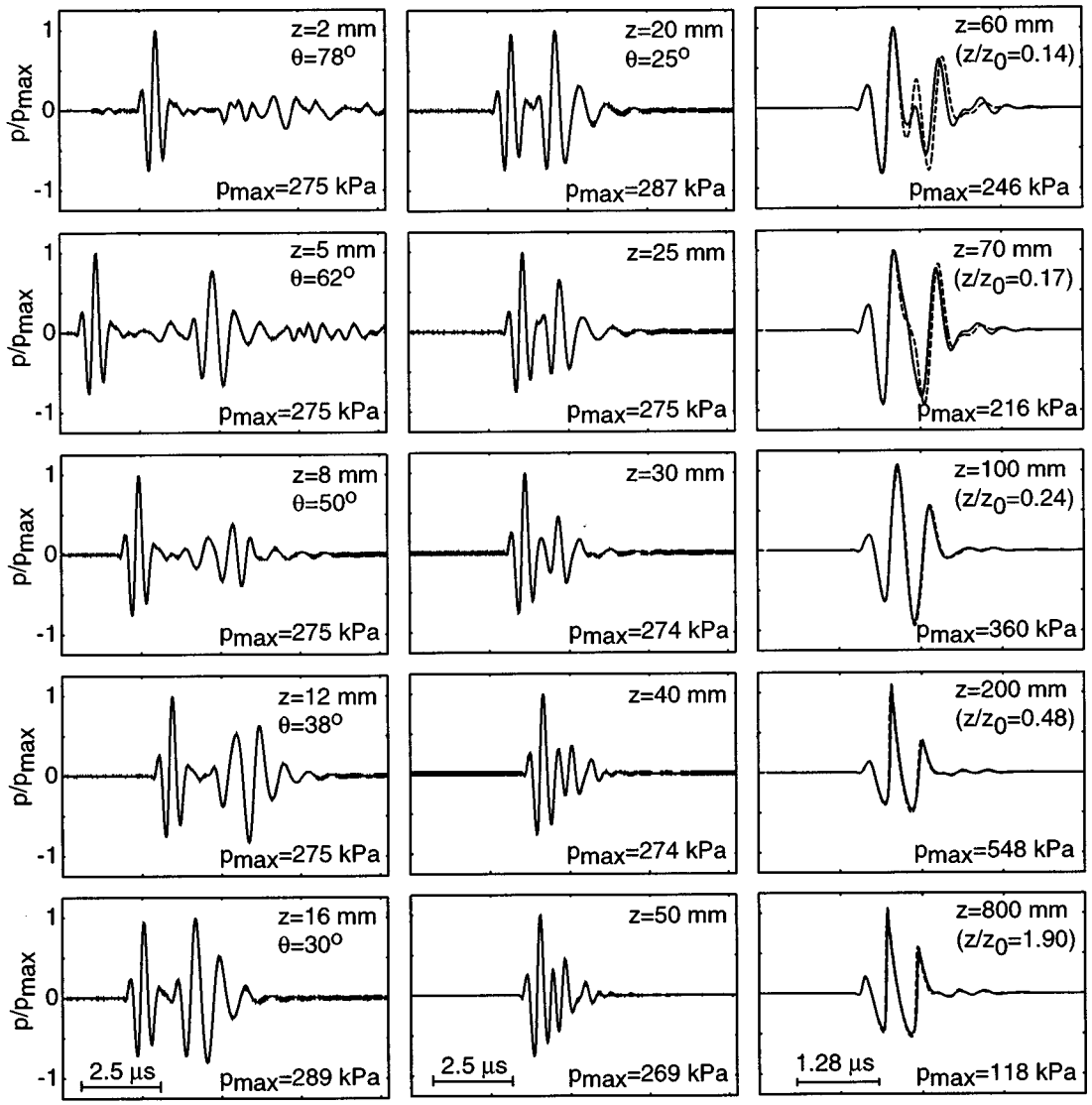


FIG. 4. Measurements (solid lines; $f_0=2.25$ MHz, $a=9.4$ mm, $z_0=420$ mm) of a short pulse along the axis of a plane piston, and comparison with theory (dashed lines in third column; $A_u=0.053$, $N_u=1.72$).

Fig. 4. In particular, subsequent measurements[24] show that at 2 MHz the hydrophone used in our experiments has a full-width half-angle, where the first null appears, of approximately 43°.

The second and third columns in Fig. 4 show the coalescence of the direct and edge waves in the region from $z=25$ mm to $z=100$ mm, where finite amplitude effects on the wave remain minor. Between $z=100$ mm and $z=200$ mm the effect of nonlinearity becomes pronounced, and a shock is formed in the center of the pulse. The distortion increases further out to $z=800$ mm.

In the third column of Fig. 4 we compare measurements with theoretical predictions based on the time domain code[17] for solving the KZK equation (dashed lines). We have expanded the time axis by a factor of 2 in order to better see the comparisons between theory and experiment. The source condition used as an input to the numerical algorithm is the measured direct wave at $z=2$ mm in Fig. 4 (we deleted the edge wave). The small amount of ringing associated with turning the source on and off is also included, and we expect to see this effect also propagate as a part of our solution. Based on the source parameters ($f_0=2.25$ MHz, $a=9.4$ mm, and $p_0=274$ kPa) and the properties of fresh water ($\rho_0=998$ kg/m^3, $c_0=1486$ m/s, and $\beta=3.5$), the absorption and nonlinearity parameters are $N_u=1.72$ and $A_u=0.053$. We start the comparison at range $z/z_0=0.14$ ($z=60$ mm), and not any closer to the source because the theoretical model becomes increasingly inaccurate. Specifically, for a circular piston the parabolic approximation is valid on axis for[25] $z \gtrsim a(ka)^{1/3}$, which corresponds to $z \gtrsim 42$ mm for the parameters of our experiment. Errors introduced on axis by the parabolic approximation are associated with the arrival time of the edge wave. At $z/z_0=0.14$ the agreement between measurement and theory is good for the center wave, but less so for the edge wave. As we move further out, the edge wave catches up with the center wave, and at about $z/z_0=0.24$ the two signals overlap. After that range the wave distorts nonlinearly. The overall agreement between the mea-

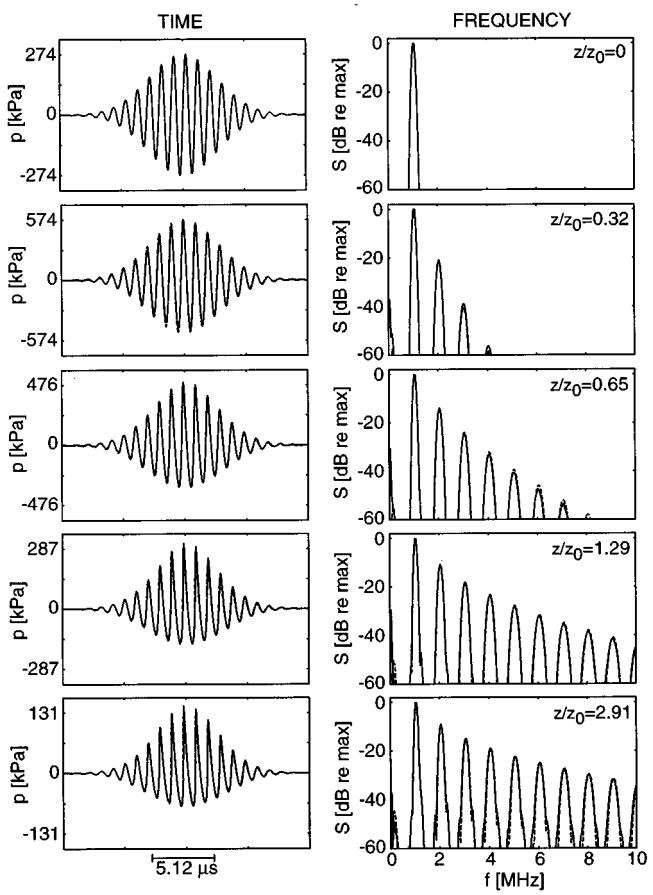


FIG. 5. Comparisons of measurements (solid lines; $f_0=1$ MHz, $a=12.1$ mm, $z_0=310$ mm) of a tone burst with Gaussian envelope ($\omega_0 T=16\pi$, $m=1$) along the axis of a plane piston with theory (dashed lines; $A_u=0.007$, $N_u=0.57$).

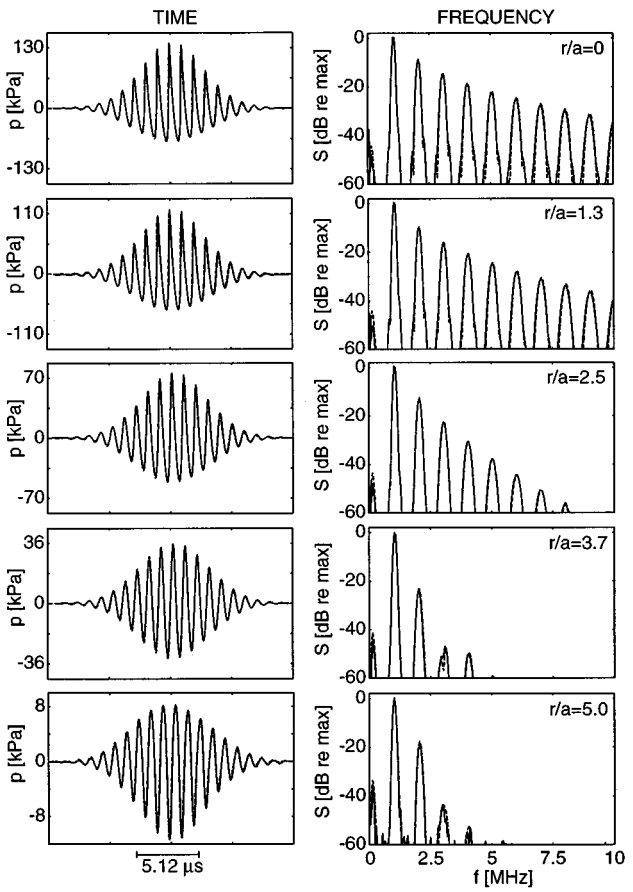


FIG. 6. Comparisons of measurements (solid lines) with theory (dashed lines) across the beam for the pulse in Fig. 5 at $z/z_0=2.91$.

surements and the theoretical predictions is very good.

We now present measurements of a Gaussian (envelope) tone burst radiated by the 1-MHz plane piston. Except for the first waveform (the source function), all measurements were made beyond the last near-field null in the propagation curve at 1 MHz. The values $\omega_0 T=16\pi$ and $m=1$ determined the relative duration and rise time of the pulse, respectively, and they defined the input functions for both the signal generator (LeCroy 9112 arbitrary function generator) and the theoretical calculations shown as dashed lines in Fig. 5. The absorption and nonlinearity parameters are $A_u=0.007$ and $N_u=0.57$, respectively. The left column in Fig. 5 contains the measured and calculated pressure waveforms $p(t')$, and the right column shows the corresponding spectral magnitudes S, which are obtained with a fast Fourier transform (FFT) of the time waveforms (both measured and calculated). Note that absolute pressures are provided for the time waveforms, whereas the frequency spectra are normalized by their peak values at the given location. At $z/z_0=0$ (at the source), we show the source condition that was supplied to both the function generator and the numerical solution. As the pulse propagates in water, the higher harmonic bands increase in level as expected, and the time waveforms distort. The agreement between measurement and theory is excellent (in fact, indistinguishable in the waveforms, although we note that the positive waveform peaks are slightly underpredicted by the theory and the negative peaks are overpredicted by equal amounts). We were limited to a maximum distance $z/z_0=2.91$ ($z=900$ mm) because of the size of the water tank.

Figure 6 contains measurements and theoretical predictions of the Gaussian pulse across the beam at range $z/z_0=2.91$, which corresponds to the range of the last waveform shown in Fig. 5. The left column shows the time waveforms and the right column the corresponding normalized frequency spectra, for both measurements and theoretical predictions. Since the higher harmonics are generated in narrower beams, we expect to see them decay more rapidly than the fundamental as we move off axis. Indeed, that is what happens in Fig. 6 from $r/a=0$ to $r/a=3.7$, in which region the distortion of the time waveforms diminishes accordingly (i.e., the shock front disappears). We note that linear theory predicts the first null in the far-field beam pattern at 1 MHz to occur at $r/a=5.6$.

In both Figs. 5 and 6 a nonlinearly generated low-frequency band (below the fundamental component) is seen. This frequency band is associated with self-demodulation.[26] As seen in Fig. 6, the low-frequency spectrum grows relative to the harmonics of the fundamental as we move off axis, because the low-frequency components have broader directivities. The agreement between measurements and numerical predictions for the low-frequency components is not

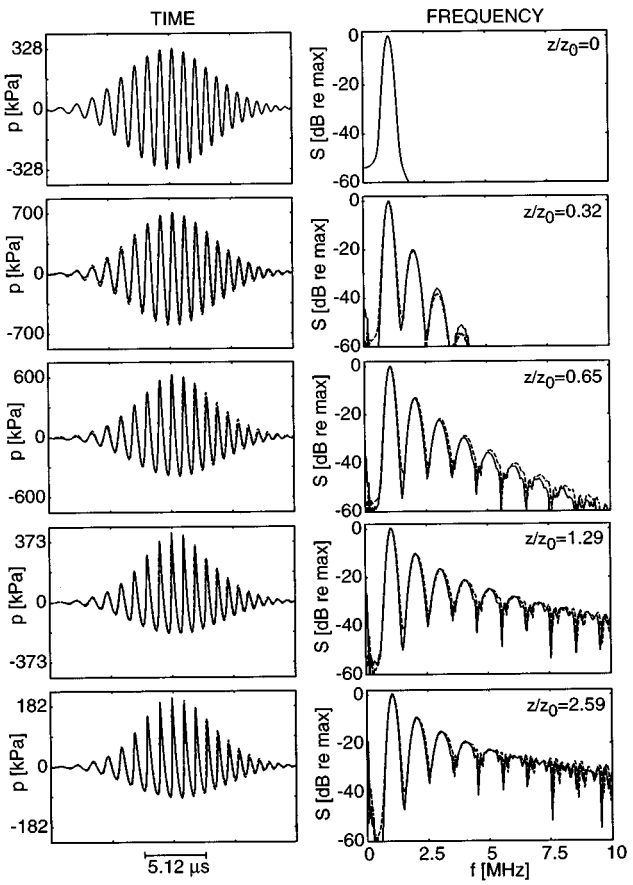


FIG. 7. Comparisons of measurements (solid lines; $f_0=1$ MHz, $a=12.1$ mm, $z_0=310$ mm) of a frequency modulated pulse having a Gaussian envelope $[\omega_0 T=16\pi,\ m=1,\ \phi(t)=(\omega_0 t)^2/120\pi]$ along the axis of a plane piston with theory (dashed lines; $A_u=0.007$, $N_u=0.684$).

quite as good as for the other components. A possible explanation is the receiver frequency response at low frequencies. It is known that the response of this type of hydrophone departs from ideal as the frequency is lowered below 1 MHz. Self-demodulation has been investigated in detail elsewhere,[27] with both theory and experiment, for the case of strong absorption.

Shown in Fig. 7 are axial measurements and theoretical predictions for a frequency-modulated Gaussian pulse with a center frequency of $f_0=1$ MHz and a phase modulation defined by $\phi(t)=(\omega_0 t)^2/120\pi$. The remaining source parameters are $\omega_0 T=20\pi$, $m=1$, $A_u=0.007$, and $N_u=0.684$. The instantaneous angular frequency of the tone is $\Omega/2\pi=(1+4f_0 t/120)f_0$, which increases linearly with time by approximately 50% over the duration of the pulse. As in Fig. 5, the waveform at $z=0$ represents the electrical input to the transducer, as well as the numerical input to the computer code. The general trends in Fig. 7 are the same as those in Fig. 5. However, the spectral bands are now wider because of the broader frequency band at the source, and at $z/z_0=2.59$ the high-frequency spectrum is very noiselike in appearance. The agreement between measurement and theory is again very good.

According to linear acoustical theory, the received sound pressure level at a fixed point increases linearly with

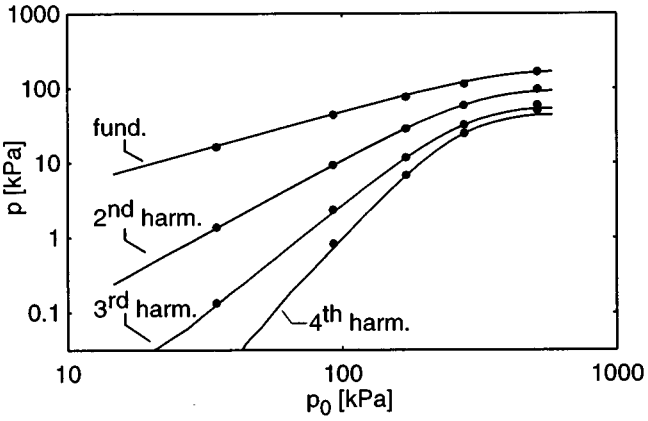


FIG. 8. Measurements (solid lines; $f_0=1$ MHz, $a=12.1$ mm, $z_0=310$ mm) and theoretical predictions (dots; $A_u=0.007$) of acoustic saturation of a finite amplitude tone burst ($\omega_0 T=34\pi$, $m=4$) in an unfocused beam at $z/z_0=2.3$.

increasing source level. At sufficiently high amplitudes, however, the amount of energy transmitted through a medium is limited by shock formation. The limit at which all additional energy supplied by the source is dissipated between the source and receiver is known as acoustic saturation. Acoustic saturation in cw fields has been investigated experimentally in the past, first for plane and spherical waves,[28,29] and subsequently for directional sound beams.[30,31] In Fig. 8 we compare theory and experiment for acoustic saturation associated with a pulsed sound beam.

With the source function defined by $\omega_0 T=34\pi$ and $m=4$, the amplitude of the 1-MHz transducer was increased while the axial sound pressure at $z=700$ mm ($z/z_0=2.3$) was monitored with the hydrophone. The peak amplitudes of the lowest four frequency bands were then plotted, solid lines for measurements and circles for theory. For each source level at which we ran the computer code, we took an additional three measurements along the axis of the source and compared them with theory to ensure that the complete field was accurately modeled along the entire propagation path. The agreement between experiment and theory was equally good at all amplitudes. Acoustic saturation is one case in which experiments are easier to perform than are the calculations. For reference, in Fig. 5 the source pressure was 274 kPa, and it corresponds to the fourth (from left) theoretical point in Fig. 8, which marks the onset of acoustic saturation. The distance at which the measurements were made for Fig. 8 was 700 mm ($z/z_0=2.3$), which corresponds to waveform distortion at an intermediate stage between the last two waveforms of Fig. 5.

B. Focused pistons

The focused 2.25-MHz piston ($G=10.5$) described in Sec. II was used in all experiments reported in the present section.

The measured waveforms in Fig. 9 were created by using the parameters $\omega_0 T=14\pi$ and $m=2$ to define the input function for the signal generator. The source waveform used in the theoretical calculations is the measured direct wave at

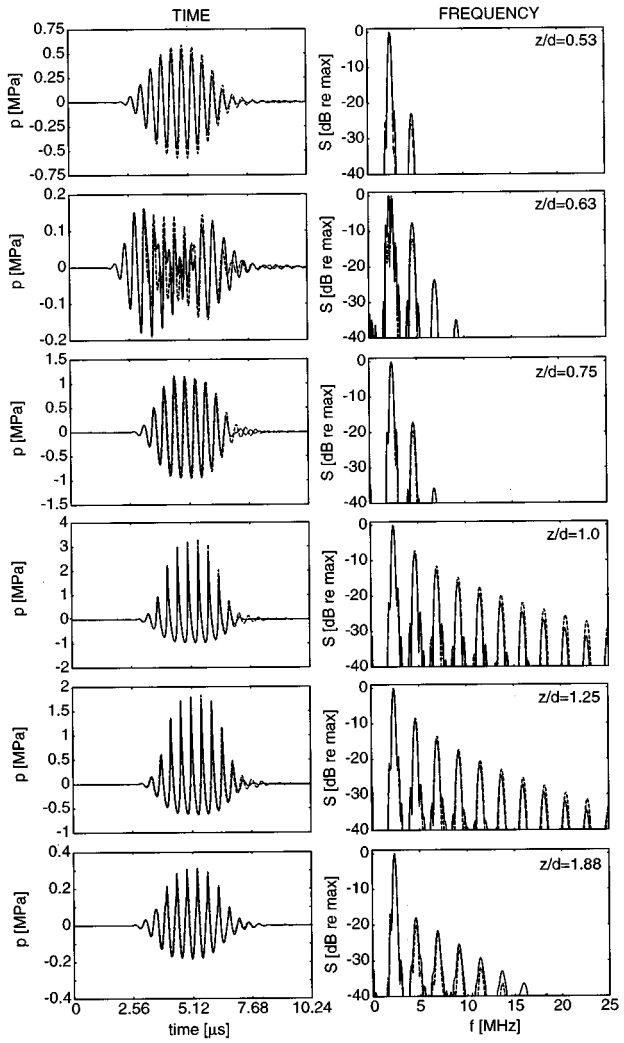


FIG. 9. Comparisons of measurements (solid lines; $f_0=2.25$ MHz, $a=18.8$ mm, $d=160$ mm) of a tone burst ($\omega_0 T=14\pi$, $m=2$) with theory (dashed lines; $A_f=0.025$, $N_f=0.34$, $G=10.5$) along the axis of a focused piston.

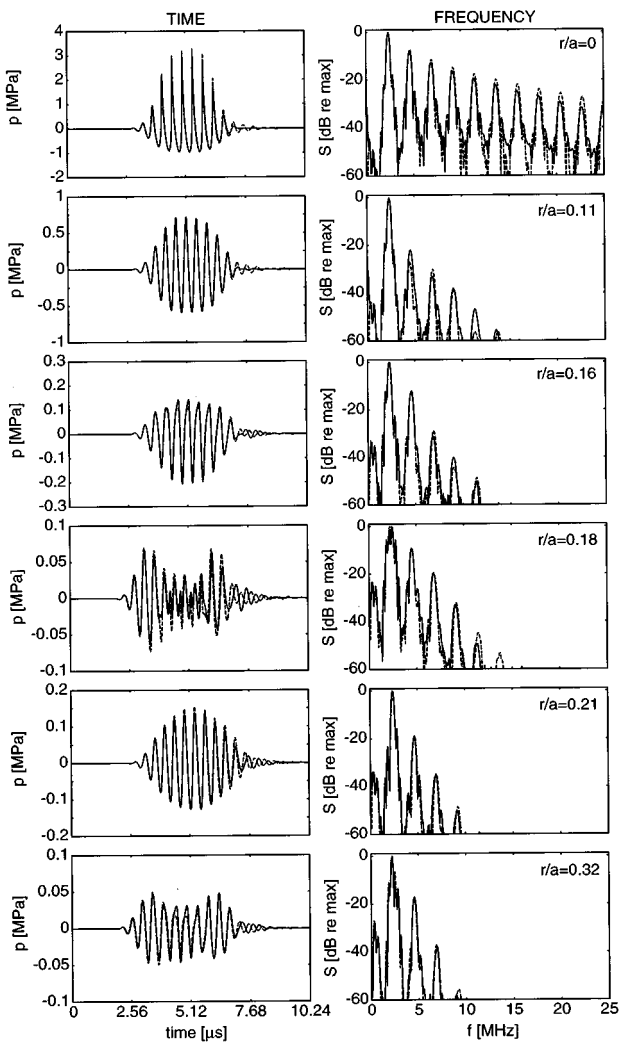


FIG. 10. Comparisons of measurements (solid lines) with theory (dashed lines) across the beam for the pulse in Fig. 9 at $z/d=1.0$.

$z=10$ mm ($z/d=0.06$, not shown). For a focused circular piston, the pressure field along the axis is also composed of two replicas of the incident waveform, as with the plane piston (see discussion of Fig. 3). However, the two contributions arrive at the same time only at the focus. In the prefocal region the direct wave precedes the diffracted wave, whereas in the postfocal region the diffracted wave arrives first. At $z=10$ mm the center and edge waves were separated completely by about three cycles. The absorption and nonlinearity parameters are $A_f=0.025$ and $N_f=0.34$, respectively.

The left column in Fig. 9 contains the measured and calculated pressure waveforms $p(t')$, and the right column shows the corresponding normalized spectral magnitudes S obtained by taking the FFT of $p(t')$. The range $z/d=0.53$ corresponds to a prefocal maximum in the propagation curve at 2.25 MHz, where the waveform closely resembles the signal at the source. The range $z/d=0.63$ corresponds to the last prefocal null, where the direct wave and edge wave differ by almost 180° and thus cancel in the center of the pulse. The apparent increase in the duration of the waveform at $z/d=0.63$ in comparison with the waveforms presented just above and below is due in part to the magnified scale that was used to compensate for the reduction in amplitude due to phase cancellation. As the pulse propagates towards the focus the amplitude increases, the higher harmonic bands increase in level, and the time waveforms distort. Beyond the focus the waveforms decrease in amplitude and become less distorted. At the focus, the agreement of theory with experiment worsens with increasing frequency. The difference between experiment and theory is less than 1 dB for the first four harmonics and it increases to about 3 dB at the tenth harmonic. The increasing discrepancies at the higher harmonics may be attributed to spatial averaging because the beamwidths become comparable to the diameter of the active element of the hydrophone.

Figure 10 contains measurements (solid lines) and theoretical predictions (dashed lines) of a pulse with the same parameters as in Fig. 9, at positions across the beam in the focal plane ($z/d=1.0$). The left column shows the time waveforms and the right column the corresponding normalized frequency spectra. The positions $r/a=0.18$ and $r/a=0.32$ correspond to nulls predicted by linear theory at 2.25

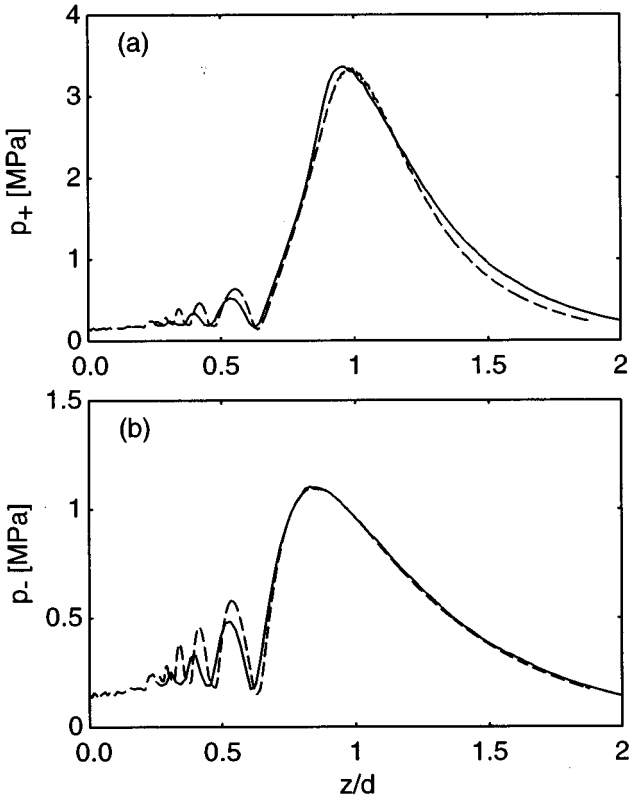


FIG. 11. Comparison of measurements (solid lines; $f_0=2.25$ MHz, $a=18.8$ mm, $d=160$ mm) and theory (dashed lines; $A_f=0.025$, $N_f=0.34$, $G=10.5$) of the axial peak positive (a) and negative pressures (b) of the pulse in Fig. 9.

MHz, the center frequency of the source waveform. Note also the nonlinearly generated low-frequency band, below the primary frequency band. Low-pass filtering of the waveforms shown in Fig. 10 reveals a waveform that resembles the second derivative of the square of the pulse envelope.

Axial propagation curves of the peak positive (p_+) and negative (p_-) pressures for the pulse shown in Fig. 9 are shown in Fig. 11. The measurements are shown as solid lines and the theoretical predictions as dashed lines. We note that the maximum value of p_+ occurs very close to the geometric focus of the source ($z/d\simeq 1.0$). However, p_- is maximized earlier ($z/d\simeq 0.8$), the significance of which is that the negative pressure plays an important role in acoustic cavitation. In medical ultrasound applications where cavitation activity may be deleterious, special care must be given to monitor cavitation in the prefocal region. *In vitro* and *in vivo* experiments[32,33] that show increased cavitation activity in the prefocal region of a lithotripter are in agreement with our results. The numerical model described here has been applied to lithotripsy fields[34,35] as well, where higher amplitudes and broader frequency pulses are utilized and similar observations have been noted.

The comparisons shown in Fig. 11 were repeated for a much shorter pulse (about 2 cycles), and these results are presented in Fig. 12. Figure 12(a) shows the measured waveform and the theoretical prediction at the focus. Comparisons of measurements and theory for the propagation curves of

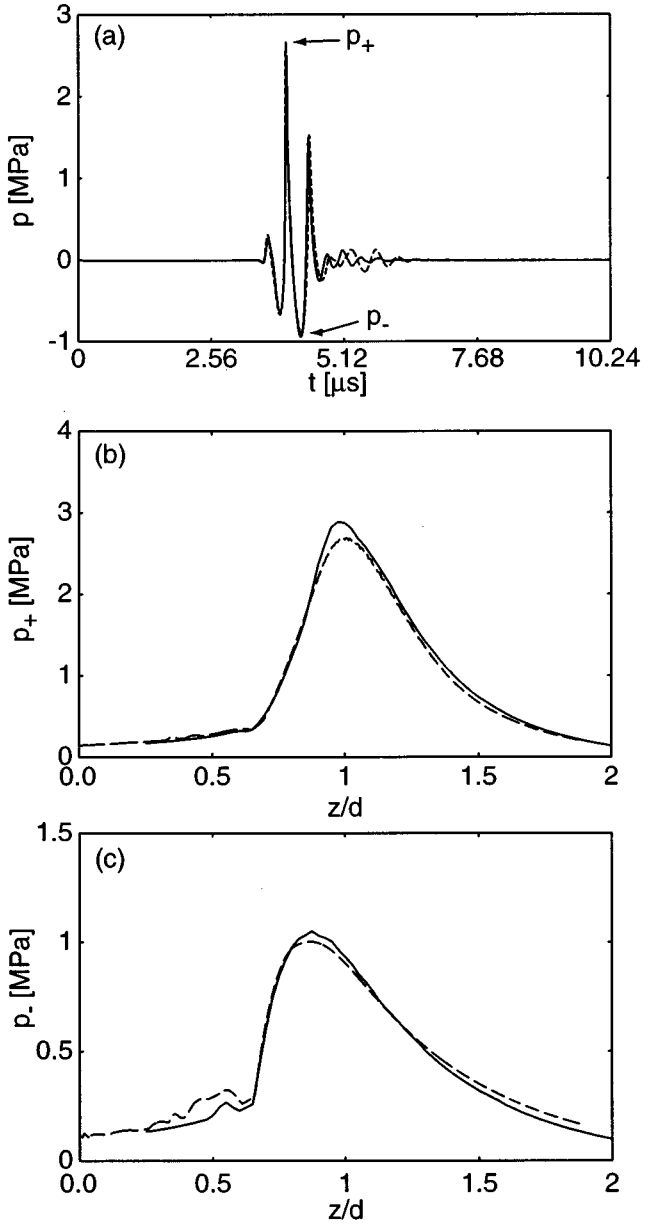


FIG. 12. Comparison of measurements (solid lines; $f_0=2.25$ MHz, $a=18.8$ mm, $d=160$ mm) and theory (dashed lines; $A_f=0.025$, $N_f=0.35$, $G=10.5$) for the axial waveform at the focus (a) and the propagation curves of the peak positive (b) and negative pressures (c) of a short pulse.

p_+ and p_- are shown in Fig. 12(b) and (c), respectively. Note that the negative pressure is maximized in the prefocal region, whereas the positive pressure is maximized closer to the geometric focus. A difference between Figs. 11 and 12 is that the near-field diffraction oscillations in Fig. 11 are not seen in Fig. 12 because of the shorter pulse length.

Shown in Fig. 13 are propagation curves for p_+ and p_- with $p_0=28.4$, 85.2, 142, and 189 kPa. The values $\omega_0 T=14\pi$ and $m=2$ determined the relative duration and rise time of the pulse (same source function parameters as in Fig. 9). The peak positive pressure is maximized close to $z/d=1$, with the exception of the case with $p_0=189$ kPa, for which there is a prefocal shift. A very slight prefocal shift is also seen for the 142 MPa pressure. This shift may be due to

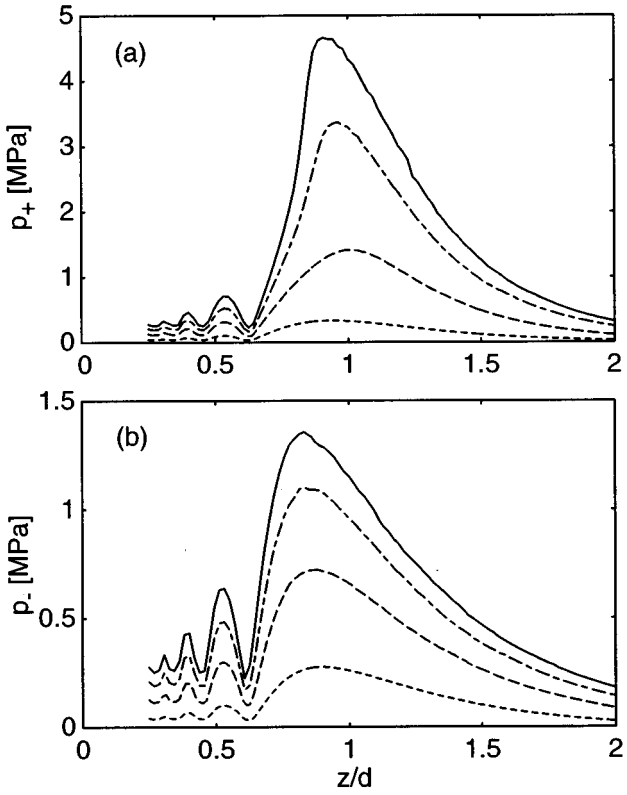


FIG. 13. Measurements of the peak positive (p_+) and negative pressures (p_-) along the axis of a focused piston ($f_0=2.25$ MHz, $a=18.8$ mm, $d=160$ mm) as function of source pressure: $p_0=28.4$, 85.2, 142, and 189 kPa, in bottom to top order.

bandwidth limitations of the hydrophone which, for the large pressure amplitudes involved ($p_+>3$ MPa), introduce uncertainty in the observed spikes associated with the peak shock pressures. Since the corresponding waveforms contained strong shocks and thus had a frequency content that extended beyond the frequency response of the hydrophone, they contained a fair amount of ringing. The maximum value of p_- is located clearly before the focus and shifts closer to the source with increasing source pressure. Errors due to hydrophone frequency response are not easily observed with measurements of p_- because the rarefaction phase of the waveform is less sensitive to measurement bandwidth than the compression phase.

IV. CONCLUSION

Detailed measurements of pulses radiated by plane and focused circular pistons in water have been shown. Emphasis was placed on nonlinear distortion of pulses. The measurements exhibit high spatial resolution and extend over a dynamic range of up to 80 dB. Thus, they enabled very precise comparisons to be made with numerical calculations based on the KZK equation. In particular, detailed comparisons of theory and experiment were made of waveforms and their corresponding spectra. The predictions are based on a time-domain algorithm for solving the KZK equation. This time-domain approach is particularly suitable for modeling pulses, and short transients in general. Acoustic saturation of a pulse was investigated. Axial propagation curves of the peak negative pressure (p_-) of an ultrasonic pulse radiated by a focused piston source indicate that the maximum negative pressure occurs in the prefocal region, a result that is relevant to acoustic cavitation studies. Finally, increasing the source pressure results in shifting of the maximum p_- closer to the source.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research, the David and Lucile Packard Foundation Fellowship for Science and Engineering, and the National Institutes of Health (Grant No. DK-43881).

[1] A. J. Coleman and J. E. Saunders, ''A review of the physical properties and biological effects of the high amplitude acoustic fields used in extracorporeal lithotripsy,'' Ultrasonics **31**, 75–89 (1993).
[2] D. A. Hutchins and G. Hayward, ''Radiated fields of ultrasonic transducers,'' in *Ultrasonic Measurement Methods*, Vol. 19 of *Physical Acoustics*, edited by R. N. Thurston and A. D. Pierce (Academic, New York, 1990), pp. 1–80.
[3] M. M. Goodsitt, E. L. Madsen, and J. A. Zagzebski, ''Field patterns of pulsed, focused, ultrasonic radiators in attenuating and nonattenuating media,'' J. Acoust. Soc. Am. **71**, 318–329 (1982).
[4] H. Djelouah, J. C. Baboux, and M. Perdix, ''The transient field of a planar ultrasonic transducer coupled to a lens: Experiments and simulations,'' J. Acoust. Soc. Am. **87**, 76–80 (1990).
[5] H. Djelouah, J. C. Baboux, and M. Perdix, ''Theoretical and experimental study of the field radiated by ultrasonic focused transducers,'' Ultrasonics **29**, 188–200 (1991).
[6] A. C. Baker and V. F. Humphrey, ''Distortion and high-frequency generation due to nonlinear propagation of short ultrasonic pulses from a plane circular piston,'' J. Acoust. Soc. Am. **92**, 1699–1705 (1992).
[7] A. C. Baker and V. F. Humphrey, ''Nonlinear propagation of short ultrasonic pulses in focused fields,'' in *Frontiers of Nonlinear Acoustics: Proceedings of 12th ISNA*, edited by M. F. Hamilton and D. T. Blackstock (Elsevier Applied Science, London, 1990), pp. 185–190.
[8] A. C. Baker and V. F. Humphrey, ''Nonlinear propagation in pulsed ultrasonic transducers,'' in *Proceedings of Ultrasonics International 89* (Butterworth, London, 1989), pp. 691–696.
[9] E. A. Zabolotskaya and R. V. Khokhlov, ''Quasi-plane waves in the nonlinear acoustics of confined beams,'' Sov. Phys. Acoust. **15**, 35–40 (1969).
[10] V. P. Kuznetsov, ''Equation of nonlinear acoustics,'' Sov. Phys. Acoust. **16**, 467–470 (1970).
[11] R. E. Apfel and C. K. Holland, ''Gauging the likelihood of cavitation from short-pulse, low duty cycle diagnostic ultrasound,'' Ultrasound Med. Biol. **17**, 179–185 (1991).
[12] J. Lighthill, *Waves in Fluids* (Cambridge U.P., Cambridge, 1980), pp. 78–83.
[13] D. T. Blackstock, ''Nonlinear Acoustics (Theoretical),'' in *American Institute of Physics Handbook* (McGraw-Hill, New York, 1972), 3rd ed., Chap. 3n.
[14] E. H. Vefring, J. Naze Tjøtta, and S. Tjøtta, ''Effects of focusing on the nonlinear interaction between two collinear finite amplitude sound beams,'' J. Acoust. Soc. Am. **89**, 1017–1027 (1991).
[15] J. Naze Tjøtta, S. Tjøtta, and E. H. Vefring, ''Propagation and interaction of two collimated finite amplitude sound beams,'' J. Acoust. Soc. Am. **88**, 2859–2870 (1990).
[16] J. Naze Tjøtta and S. Tjøtta, ''Nonlinear equations of acoustics, with application to parametric acoustic arrays,'' J. Acoust. Soc. Am. **69**, 1644–1652 (1981).
[17] Y. S. Lee and M. F. Hamilton, ''Time-domain modeling of pulsed finite-amplitude sound beams,'' J. Acoust. Soc. Am. **97**, 906–917 (1995).
[18] R. O. Cleveland, M. F. Hamilton, and D. T. Blackstock, ''Time-domain modeling of finite-amplitude sound in relaxing fluids,'' J. Acoust. Soc. Am. **99**, 3312–3318 (1996).
[19] Y.-S. Lee, ''Numerical solution of the KZK equation for pulsed finite amplitude sound beams in thermoviscous fluids,'' Ph.D. dissertation, The University of Texas at Austin, 1993.

[20] J. A. TenCate, ''An experimental investigation of the nonlinear pressure field produced by a plane circular piston,'' J. Acoust. Soc. Am. **94**, 1084–1089 (1993).
[21] M. A. Averkiou, ''Experimental investigation of propagation and reflection phenomena in finite-amplitude sound beams,'' Ph.D. dissertation, The University of Texas at Austin, 1994, Appendix A.
[22] M. A. Averkiou and M. F. Hamilton, ''Measurements of harmonic generation in a focused finite-amplitude sound beam,'' J. Acoust. Soc. Am. **98**, 3439–3442 (1995).
[23] D. R. Bacon, ''Characteristics of a pvdf membrane hydrophone for use in the range 1–100 MHz,'' IEEE Trans. Sonics Ultrason. **SU-29**, 18–25 (1982).
[24] B. J. Landsberger, ''Second harmonic generation in sound beams reflected from and transmitted through immersed elastic solids,'' Ph.D. dissertation, The University of Texas at Austin, 1997, Appendix C.
[25] J. Naze Tjøtta and S. Tjøtta, ''An analytical model for the nearfield of a baffled piston transducer,'' J. Acoust. Soc. Am. **68**, 334–339 (1980).
[26] H. O. Berktay, ''Possible exploitation of non-linear acoustics in underwater transmitting applications,'' J. Sound Vib. **2**, 435–461 (1965).
[27] M. A. Averkiou, Y.-S. Lee, and M. F. Hamilton, ''Self-demodulation of amplitude- and frequency-modulated pulses in a thermoviscous fluid,'' J. Acoust. Soc. Am. **94**, 2876–2883 (1993).
[28] C. H. Allen, ''Finite amplitude distortion in a spherically diverging sound wave in air,'' Ph.D. dissertation, Pennsylvania State University, 1950.
[29] D. A. Webster and D. T. Blackstock, ''Finite-amplitude saturation of plane sound waves in air,'' J. Acoust. Soc. Am. **62**, 518–523 (1977).
[30] J. A. Shooter, T. G. Muir, and D. T. Blackstock, ''Acoustic saturation of spherical waves in water,'' J. Acoust. Soc. Am. **55**, 54–62 (1974).
[31] T. L. Riley, ''Generation of harmonics in finite amplitude sound radiated in water by a circular piston,'' M.S. thesis, The University of Texas at Austin, 1983.
[32] A. J. Coleman, M. Whitlock, T. Leighton, and J. E. Saunders, ''The spatial distribution of cavitation induced acoustic emission, sonoluminescence and cell lysis in the field of a shock wave lithotripter,'' Phys. Med. Biol. **38**, 1545–1560 (1993).
[33] A. J. Coleman, T. Kadama, M. J. Choi, T. Adams, and J. E. Saunders, ''The cavitation threshold of human tissue exposed to 0.2 MHz pulsed ultrasound: Preliminary measurements based on a study of clinical lithotripsy,'' Ultrasound Med. Biol. **21**, 405–417 (1995).
[34] M. A. Averkiou, L. A. Crum, and M. F. Hamilton, ''Theoretical modeling of the acoustic pressure field produced by commercial lithotripters,'' J. Acoust. Soc. Am. **98**, 2941(A) (1995).
[35] M. A. Averkiou and L. A. Crum, ''Cavitation: Its role in stone comminution and renal injury,'' in *Topics in Clinical Urology: New Developments in the Management of Urolithiasis*, edited by J. Lingeman and G. M. Preminger (Igaku-Shoin Medical Publishers, New York, 1995), pp. 21–40.

Side drift of aerosols in two-dimensional resonant acoustic levitators

Y. Dain[a)]
Faculty of Mechanical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel

(Received 14 February 1996; revised 6 May 1997; accepted 14 July 1997)

The periodic structure of dispersed particle distributions in a quickly oscillating plane sonic field has been the object of many classical investigations. It is well-known that particles dispersed in a standing one-dimensional sonic wave tend to concentrate either at nodes or at anti-nodes of the fluid velocity. If the amplitude of oscillations is sufficiently small, such that no shock waves can develop, the interaction of a two-dimensional sonic mode with particles in a resonant chamber is analyzed. The chamber considered is bounded by two parallel cylindrical mirrors, which enable the existence of the ''bouncing ball'' mode concentrated in the neighborhood of the minimal diameter connecting the mirrors centers. If the size of the diameter exceeds the sonic wavelength, the mode magnitude decays rapidly away from the diameter. The domain is filled with fluid, which contains spherical particles. The quantitative theoretical description of the phenomenon of steady forces acting on dispersed particles in the chamber is achieved. The chamber boundary mirrors cause dispersed particles to move in a direction transverse to the diameter, where the ''bouncing ball'' mode is concentrated. It is shown that in addition to the well-known one-dimensional particle motion along the diameter toward the velocity nodes or anti-nodes, there exists a side drift, which depends on the fluid-to-particle density ratio, the ''bouncing ball'' mode frequency and the radii of curvature of the resonator's mirrors. © *1997 Acoustical Society of America.* [S0001-4966(97)05911-0]

PACS numbers: 43.25.Uv [MAB]

INTRODUCTION

The levitation of objects by sound fields is one of the interesting manifestations of nonlinear acoustics. An application of this phenomenon of some technological and economic importance is in material positioning on containerless processing. Such positioning is regarded as the ability to localize a sample at a position interior to the chamber boundaries, where aerosols and suspensions will not have been subjected to the contaminating influence of a solid chamber boundary. Acoustic levitation also creates a useful environment for the investigation of interesting physical phenomena. It is possible to envisage fascinating experiments to probe the static and dynamic properties of a multiphase flow.

It is well-known that aerosol particles in a standing one-dimensional sonic wave tend to concentrate either at nodes or at antinodes of the fluid velocity field. In the course of the interaction of plane standing sound waves with particles, hydrodynamic forces produce a steady particle motion, which is characterized by an averaged force. This so-called drift force depends on particle size, frequency of oscillations, and fluid-to-particle density ratio.

A drift force acting on small particles is of a nature related to that of the radiation pressure forces in a sound field; it was studied by various authors for small spheres placed freely in a plane stationary wave field. An expression for the drift force for an ideal compressible fluid model was first obtained by King (1934). Calculations were given when particles are small with respect to the acoustic wavelength and the Stokes friction force is negligible. A more exact expression for the drift force was derived by Westervelt (1951). The acoustic radiation pressure on spheres of finite compressibility placed in a plane sound wave was calculated by Gorkov (1962) and Yosioka and Kawasima (1955).

Aerosol particle motion in the plane standing sound wave under the action of the Stokes friction force was studied by Duhin (1960). More recently, the effect of particle concentration in a plane standing sound wave was studied by Czyz (1987, 1990), Fittipaldi (1979), Hager and Benes (1991), Higashitani (1981), Nigmatulin (1990), Rudnick (1951), Vainshtein *et al.* (1992), and Westervelt (1950, 1957).

The drift force acting on the particle in a spherically converging or diverging wave was obtained by Gorkov in the case when the Stokes friction force is negligible (Gorkov, 1962). Recently the analysis of the aerosol motion in two-dimensional standing sound waves under the action of the Stokes friction force was performed by Dain *et al.* (1994, 1995). The motion of aerosol particles has been studied in the two-dimensional vessel bounded by a plane vibrating wall and a cylindrical stationary mirror with a small curvature which is responsible for the distortion of the wave front. In Dain *et al.* (1994) the region of the vessel was supposed to be much smaller; in Dain *et al.* (1995) the region of the vessel was supposed to be much bigger than the sonic wavelength. The effect of additional sidewise drift force in the vicinity of the minimum region diameter has been observed. It was shown that a small distortion of the plane stationary wave results in a sidewise particle scattering away from the vicinity of the diameter or a convergent motion toward it. In the experiments of Trinh and Robey (1994) eddies at the velocity anti-nodes along the axis of symmetry have been observed in resonant chamber with concave axially symmet-

[a)]Present address: Department of Physics, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK.

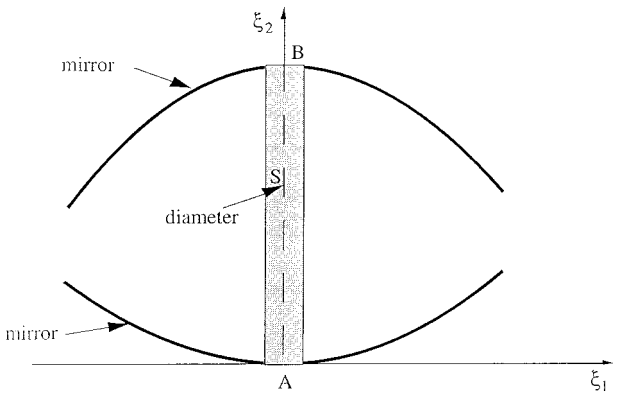


FIG. 1. Schematic representation of the core region of the resonant chamber.

ric reflector. Motivation for the present investigation comes from the need to understand in detail the acoustic localization capabilities of various resonator chamber geometry. The resonator chambers used may have rather complicated physical boundaries. As a result, the acoustic field is not readily described as either a plane standing wave or as a single mode of an ideal rectangular chamber. The main emphasis, therefore, is to determine those normal modes that will sustain stable acoustic positioning of small samples, such as the rigid phase in dilute aerosols and suspensions. An additional advantage of such modes is that it is possible to generate an intense sound field by driving them at, or close to, their resonant frequency by means of a typical acoustic levitator in which one of the mirrors vibrates (Trinh and Robey, 1994) or some other sound wave generator placed in resonant chamber.

In this paper we report the results of an investigation of the mechanism of particle motion in the closed resonant chamber (Fig. 1), supposing that the chamber supports the so-called ''bouncing ball'' mode (confined Gaussian beam). We use the short wave asymptotic expression for such a mode, which is obtained by the parabolic equation method in Babich and Buldyrev (1991) or Buldyrev (1966). The ''bouncing ball'' mode is concentrated in a small neighborhood of a thin region centered around a minimal diameter and decreases exponentially outside of it. Therefore, our results are limited to the region where the radiation pressure is concentrated. The two-dimensional steady-state equations governing the mean particle motion are obtained. It is shown that in addition to the well-known one-dimensional particle motion along the diameter, toward node or anti-nodes of the fluid velocity, there is a sidewise particle drift. The small samples, such as a rigid phase in aerosols and suspensions, which are more dense than the host fluid, are repelled from the minimal diameter of the region [see Fig. 2(a),(b)]. Due to the purification effect, these results imply that it is not possible to agglomerate aerosols under conditions assumed. The rigid particles, which are less dense than the host fluid [see Fig. 2(c)] are supposedly incompressible. The behavior of compressible particles or bubbles trapped in a liquid is out of the scope of present investigation.

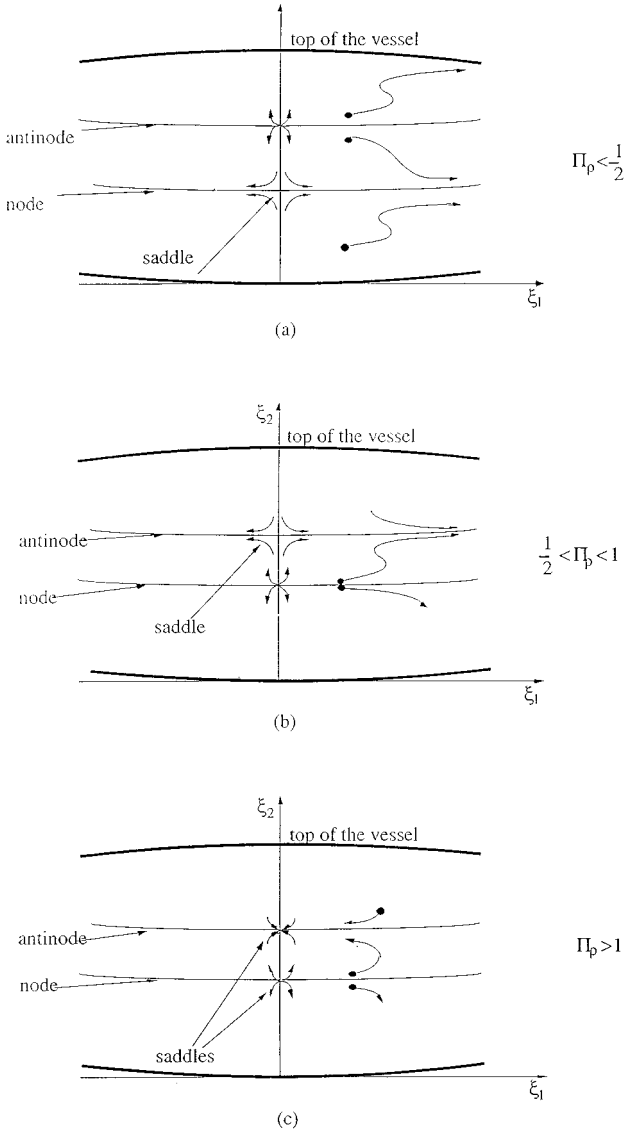


FIG. 2. Schematic description of the particle eddies in the convex resonant chamber.

I. FORMULATION OF THE PROBLEM

Consider a two-dimensional acoustic resonator, the interior of which is filled with a monodispersed suspension. The dispersed particles are assumed to be spherical and much smaller than any characteristic length of the system. The latter is a basic assumption in the analysis of multiphase flow mechanics (Nigmatulin, 1990).

Suppose that diameter S is orthogonal to the boundary Σ of resonator chamber and intersects Σ at points A and B (see Fig. 1). Let us choose axis y' to coincide with the diameter S. Then, if the origin is placed at point A, the equations of the boundary in the neighborhood of points A and B can be taken in the form

$$y' = f(x'), \quad y' = h + g(x'), \tag{1}$$

respectively, where h is the length of diameter S, and expansions of the functions $f(x')$ and $g(x')$ in the Taylor series are

$$f(x')=\frac{1}{2r_1'}x'^2,\quad g(x')=-\frac{1}{2r_2'}x'^2. \tag{2}$$

Here r_1, r_2 are the radii of the boundary curvature in points A and B. We suppose that added mass and Basset forces are negligible, in comparison with the Stokes force, and thus are excluded from the present analysis. We shall confine ourselves to a system of differential equations pertinent to the nonstationary flow of a monodisperse collision-free mixture with barotropic phases. The equations for the two-phase mixture are (see Nigmatulin, 1990):

$$\begin{aligned}
&\frac{\partial}{\partial t'}\rho_g^0\alpha_g+\nabla\cdot(\rho_g^0\alpha_g\mathbf{v}_g')=0,\\
&\rho_g^0\alpha_g\frac{d\mathbf{v}_g'}{dt'}+\alpha_g\nabla\rho'=-\alpha_p\frac{18\mu}{d^2}(\mathbf{v}_g'-\mathbf{v}_p'),\\
&\frac{\partial}{\partial t'}\rho_p^0\alpha_p+\nabla\cdot(\rho_p^0\alpha_p\mathbf{v}_p')=0,\\
&\rho_p^0\alpha_p\frac{d\mathbf{v}_p'}{dt'}+\alpha_p\nabla\rho'=\alpha_p\frac{18\mu}{d^2}(\mathbf{v}_g'-\mathbf{v}_p'),\\
&a_g+\alpha_p=1,
\end{aligned} \tag{3}$$

where α_g,α_p are the volume fractions of the fluid and particle phases, respectively, ρ_g^0,ρ_p^0 are the intrinsic phase densities, d is the particle diameter, and μ is the dynamic viscosity of the fluid.

We consider the barotropic form of the equation of state, proposed by Nigmatulin (1990):

$$p'=p_0'+C_0^2(\rho_g^0-\rho^0)+\frac{(\gamma-1)C_0^2}{2\rho^0}(\rho_g^0-\rho^0)^2, \tag{4}$$

where p_0' and ρ^0 and C_0 are the fluid pressure, density, and speed of sound in the state of rest, respectively.

The ratio of the specific heats γ is the adiabatic index in the gase case, and an empirical coefficient in the liquid case ($\gamma=6$ for water), which characterizes nonlinearity of a nondisturbed carrying phase.

We will seek the solution of nonlinear Eqs. (3) and (4), the normal component of which satisfies the boundary condition

$$v_g^{n'}|_\Sigma=0 \tag{5}$$

on the boundary Σ.

The forces of the interphase interaction affecting particles are taken in the quasi-incompressible approximation (Nigmatulin, 1990). It does not lead to noticeable errors, if the characteristic length, along which the carrying phase density varies, is much larger than the particle size. The diameter of the monodisperse particles is assumed much smaller than the characteristic wavelength of the fluid oscillations

$$d\ll C_0/\omega'. \tag{6}$$

This condition makes it possible to neglect the scattering of the pressure wave from the particle surface, and to evaluate the interphase force as if the fluid is incompressible. The characteristic relaxation time t_μ of the Stokes flow past a particle satisfies the condition

$$t_\mu=\frac{\rho^0 d^2}{18\mu}\ll\frac{1}{\omega'}. \tag{7}$$

This time is needed to develop the Stokes regime of the flow over the particle. Due to condition (7), particle inertia is negligible and the Stokes friction force prevails. The particle diameter is much smaller than the viscous boundary layer thickness due to the oscillatory flow, and the pressure force could be neglected for gas aerosols, but it is significant for liquid suspensions, which will be justified in the final formulas of particle motion equations. The separation h of the mirrors is assumed to be much bigger than the wavelength

$$h\gg\frac{C_0}{\omega'}. \tag{8}$$

The Gaussian beam will be found under the condition that its width l is much smaller than the separation of the mirrors.

$$\frac{l}{h}\sim\sqrt{\frac{C_0}{\omega h}}\ll 1. \tag{9}$$

To simplify the analysis we take ρ_p^0 to be constant, and assume low volume and mass content of the solid phase:

$$\alpha_p\ll 1,\quad \alpha_p\rho_p^0\ll\alpha_g\rho_g^0. \tag{10}$$

These restrictions are justified, because typical particle concentrations for aerosols in practical applications are

$$\alpha_p/\alpha_g<0.001. \tag{11}$$

Hence, the dynamic influence of the dispersed solid phase on the fluid motion may be ignored and system (3) may be decoupled into separate equations for the fluid and for the solid phase. The dimensionless variables

$$\begin{aligned}
&1=\frac{C_0}{h}t',\quad x=\frac{x'}{h},\quad y=\frac{y'}{h},\quad p=\frac{p'}{\rho^0C_0^2},\\
&\rho=\frac{\rho_g^0}{\rho^0},\quad \mathbf{v}_g=\frac{\mathbf{v}_g'}{C_0},\quad \mathbf{v}_p=\frac{\mathbf{v}_p'}{C_0},
\end{aligned} \tag{12}$$

are advantageous for the mixture motion, where ω' is the eigenfrequency of the eigenmode to be found. Let us introduce dimensionless groups

$$\begin{aligned}
&p_0=\frac{p_0'}{\rho^0C_0^2},\quad \Pi_\rho=\frac{\rho^0}{\rho_p^0},\quad \Pi_f=\frac{18\mu}{\rho_p^0d^2\omega'},\\
&r_1=\frac{r_1'}{h},\quad r_2=\frac{r_2'}{h},
\end{aligned} \tag{13}$$

where Π_ρ is the fluid-to-particle density ratio, Π_f supposed to be large for small particles

$$\Pi_f\gg 1. \tag{14}$$

If criterion (14) is satisfied, the Stokes friction force prevails (see Nigmatulin, 1990).

Upon substitution of Eq. (12) into systems (3) and (4) with allowance for Eq. (10), one can obtain the dimensionless system for the fluid motion:

$$\frac{\partial \rho}{\partial t}+\nabla\cdot(\rho \mathbf{v}_g)=0,$$

$$\frac{d\mathbf{v}_g}{dt}+\frac{1}{\rho}\nabla p=0, \tag{15}$$

$$p=\Pi_\rho+\rho-1+\tfrac{1}{2}(\gamma-1)(\rho-1)^2,$$

and dimensionless equations describing motion of the solid phase:

$$\frac{d\mathbf{v}_p}{dt}+\Pi_f\mathbf{v}_p=-\Pi_p\nabla\rho+\Pi_f\mathbf{v}_g. \tag{16}$$

If time is sufficiently long, the particle velocity can be asymptotically represented as a series with inverse powers of Π_f due to condition (14)

$$\mathbf{v}_p=\mathbf{v}_p^{(0)}+\Pi_f^{-1}\mathbf{v}_p^{(1)}+\dots. \tag{17}$$

To confine ourselves to the first two members of series (17), we obtain for the particle velocity the equation

$$\mathbf{v}_p=\mathbf{v}_g+\Pi_f^{-1}(\Pi_\rho\rho-1)\frac{d\mathbf{v}_g}{dt}, \tag{18}$$

which will be studied together with fluid Eqs. (15).

II. ASYMPTOTIC SOLUTION OF FLUID EQUATIONS

The basic goal of this section is to obtain a nonlinear solution of the fluid equation, the main part of which is a ''bouncing ball'' eigenmode of the linearized boundary problem. We suppose that the dimensionless amplitude of the fluid velocity ϵ is sufficiently small to cause shock waves and discontinuities in the flow:

$$\epsilon\ll 1. \tag{19}$$

We consider the boundary conditions for inviscid fluid

$$v_g^n|_\Sigma=0, \tag{20}$$

where the part of the boundary Σ near the ends of the diameter S is represented by the equations

$$y=\frac{1}{2r_1}x^2+\cdots,$$
$$y=1-\frac{1}{2r_2}x^2+\cdots. \tag{21}$$

Then the solution can be sought as a series expanded in powers of the small parameter ϵ, Eq. (19)

$$\rho=1+\epsilon\rho_1+\cdots,$$
$$p=p_0+\epsilon p_1+\cdots, \tag{22}$$
$$\mathbf{v}g=\epsilon\mathbf{v}_{g1}+\cdots,$$

where the first approximation is an eigenmode of the linear boundary problem.

Let us exclude the pressure in system (15) by means of the third equation. After substitution of expansions (23) into Eqs. (15), one has a sequence of systems of linear differential equations for the functions in series (23), the first two of which are

$$\frac{\partial\rho_1}{\partial t}+\nabla\cdot\mathbf{v}_{g1}=0,\quad \frac{\partial\mathbf{v}_{g1}}{\partial t}+\nabla\rho_1=0, \tag{23}$$

and

$$\frac{\partial\rho_2}{\partial t}+\nabla\cdot\mathbf{v}_{g2}=-\nabla\cdot(\rho_1\mathbf{v}_{g1})$$
$$\frac{\partial\mathbf{v}_{g2}}{\partial t}+\nabla\rho_2=-\rho_1\frac{\partial\mathbf{v}_{g1}}{\partial t}-(\mathbf{v}_{g1}\cdot\nabla)\mathbf{v}_{g1}-\tfrac{1}{2}(\gamma-1)\nabla(\rho_1^2). \tag{24}$$

This equation on the density replaces the pressure.

From Eq. (20) it follows that the normal components of the velocity approximations in Eqs. (23) and (24) obey boundary conditions

$$v_{g1}^n|_\Sigma=0,\quad v_{g2}^n|_\Sigma=0. \tag{25}$$

In order to obtain an eigenmode of Eqs. (23), let us seek it in the form

$$\rho_1=\tilde{\rho}(x,y)\sin\psi t,$$
$$\mathbf{v}_{g1}=\tilde{\mathbf{v}}_g(x,y)\cos\omega t. \tag{26}$$

Following Babich and Buldyzev (1991) or Buldyrev (1966), one can find solution (26) of Eqs. (23), concentrated in a thin boundary layer near diameter S and satisfying conditions

$$\tilde{\rho}\to 0,\quad \tilde{\mathbf{v}}_g\to 0, \tag{27}$$

when

$$|\omega^{1/2}x|\to\infty. \tag{28}$$

Such a solution, which is called the ''bouncing ball'' mode, can be obtained as a series in powers of the small parameter ω^{-1}, under the assumption

$$0<\left(1-\frac{1}{r_1}\right)\left(1-\frac{1}{r_2}\right)<1 \tag{29}$$

of the radii of the boundary curvature at the points of intersection with diameter S. In particular, when both conditions,

$$\left(1-\frac{1}{r_1}\right)>0,\quad \left(1-\frac{1}{r_2}\right)>0, \tag{30}$$

hold, the eigenfrequency of the ''bouncing ball'' mode can be calculated by the asymptotic formula

$$\omega_k=\pi k+\frac{1}{2}\arccos\sqrt{\left(1-\frac{1}{r_1}\right)\left(1-\frac{1}{r_2}\right)}+O\left(\frac{1}{k}\right), \tag{31}$$

where k is the value of horizontal mode lines of eigenfunctions $\tilde{\rho}$ and $\tilde{v}_g$ in Eq. (26). Then, corresponding asymptotic expressions for eigenfunctions can be written in the form

$$\tilde{\rho}^k=\frac{1}{\sqrt{F(y)}}\cos\alpha(x,y)\exp\left(-\frac{\omega_k x^2}{2F^2(y)}\right)+O\left(\frac{1}{\omega_k}\right),$$

$$\tilde{v}^k_{gx}=\frac{x}{F^{5/2}(y)}\,(-F'F\,\sin\,\alpha(x,y)$$
$$-\cos\,\alpha(x,y))\exp\left(\frac{\omega_k x^2}{2F^2(y)}\right)+O\left(\frac{1}{\omega_k}\right),$$
$$\tilde{v}^k_{gy}=-\frac{1}{\sqrt{F(y)}}\,\sin\,\alpha(x,y)\exp\left(\frac{\omega_k x^2}{2F^2(y)}\right)+O\left(\frac{1}{\omega_k}\right), \tag{32}$$

where the functions $\alpha(x,y)$ and $F(y)$ in Eq. (33) can be calculated by formulae (33) and (34) through the radii of the curvature r_1, r_2:

$$\alpha(x,y)=\omega_k y-\frac{1}{2}\int_0^y\frac{d\tau}{F^2(\tau)}+\frac{F'(y)}{2F(y)}\,\omega_k x^2+O\left(\frac{1}{\omega_k}\right), \tag{33}$$

$$F(y)=\sqrt{a_{12}\left(\frac{r_1}{r_1-1}\,f_1^2-2\,f_1 f_2+\frac{r_2}{r_2-1}\,f_2^2\right)}, \tag{34}$$

where

$$f_1(y)=-\frac{y}{r_1}+1,\quad f_2(y)=\frac{y-1}{r_2}+1, \tag{35}$$

$$a_{12}=\frac{r_1 r_2\sqrt{(r_1-1)(r_2-1)}}{r_1+r_2-1}. \tag{36}$$

Let us seek the solution of Eqs. (24) in the form

$$\rho_2=\rho_2^s+\rho_2^t,\quad \mathbf{v}_{g2}=\mathbf{v}_{g2}^s+\mathbf{v}_{g2}^t, \tag{37}$$

where the first terms are steady-state functions and the second terms harmonically depend on time. Then with substitutions of Eqs. (26) and (32), the steady-state part of Eqs. (24) takes the form

$$\frac{\partial\rho_2^s}{\partial t}+\boldsymbol{\nabla}\cdot\mathbf{v}_{g2}^s=0,$$
$$\frac{\partial\mathbf{v}_{g2}^s}{\partial t}+\boldsymbol{\nabla}\rho_2^s=-\tfrac{1}{2}\boldsymbol{\nabla}(\|\tilde{\mathbf{v}}\|_g^2+\gamma\tilde{\rho}^{\,2}), \tag{38}$$

where the normal component of the velocity satisfies boundary conditions (49). The solution of system (38), which obeys boundary conditions (25), is the steady-state pressure

$$\rho_2^s=-\tfrac{1}{2}(\|\tilde{\mathbf{v}}\|_g^2+\gamma\tilde{\rho}^{\,2}), \tag{39}$$

and zero steady-state velocity

$$\mathbf{v}_{g2}^s=0. \tag{40}$$

We assume that $2\omega_k$ is not an eigenvalue of the boundary problem under discussion and equations for harmonically depending on the time part of solution (37) have no secular terms on the right-hand sides. We do not calculate such a solution, since its effect is small after averaging particle equations with respect to time and, therefore, it does not contribute to the mean motion of particles.

Since the steady-state component of the second-order approximation of the fluid velocity is zero, the secondary fluid acoustic streaming is absent. The two-dimensional acoustic radiation pressure (39) is the cause of the aerosol particle motion, as in the case of the plane standing wave.

III. ANALYSIS OF PARTICLE MOTION EQUATIONS

Let us substitute series (50) into the particle motion Eqs. (18) and study them in a small neighborhood of diameter S. After substitution of the variable x by v,

$$v=\omega^{-1/2}x. \tag{41}$$

Equations (18) of the particle motion take the form

$$\frac{dv}{dt}=\left\{\begin{array}{l}\dfrac{\omega\epsilon^2\Pi_f^{-1}}{2F^3}\left((2\Pi_\rho-1)\left(\dfrac{FF'}{2}\sin 2\alpha(v,y)\right.\right.\\ \left.\left.\qquad+\cos 2\alpha(v,y)\right)+1\right)\exp\left(-\dfrac{v^2}{2F^2(y)}\right)\\ +\left[\begin{array}{l}\epsilon(H_1(v,y)\sin t+H_2(v,y)\cos t)\\ \quad+\epsilon^2(G_1(v,y)\sin 2t+G_2(v,y)\cos 2t)\\ \quad+\epsilon^2 v_{g2x}+\epsilon^2\Pi_f^{-1}(\Pi_\rho-1)\,\dfrac{\partial v_{g2x}}{\partial t}\end{array}\right]\exp\left(-\dfrac{v^2}{2F^2(y)}\right)\end{array}\right\}v,$$

$$\frac{dy}{dt}=\frac{\omega\epsilon^2\Pi_f^{-1}}{4F}\,(2\Pi_\rho-1)\sin 2\alpha(v,y)\exp\left(-\frac{v^2}{2F^2(y)}\right)+\left[\begin{array}{l}\epsilon(H_3(v,y)\sin t+H_4(v,y)\cos t)\\ \quad+\epsilon^2(G_3(v,y)\sin 2t+G_4(v,y)\cos 2t)\\ \quad+\epsilon^2 v_{g2y}+\epsilon^2\Pi_f^{-1}(\Pi_\rho-1)\,\dfrac{\partial v_{g2y}}{\partial t}\end{array}\right]\exp\left(-\frac{v^2}{2F^2(y)}\right). \tag{42}$$

The averaging procedure of Bogolubov and Mitropolski (1961) of Eqs. (43) yields the steady-state system of nonlinear differential equations describing the mean steady-state particle motion. All harmonic terms and time derivatives on the right-hand sides of Eqs. (43) vanish under time averaging. Let us denote

$$v_1=\zeta_1+v_t,$$
$$y_1=\zeta_2+y_t, \tag{43}$$

where variables ζ_1,ζ_2 are steady-state parts of particle coordinates, and v_t,y_t depend on time. Then the independent on time terms on the right-hand side of Eq. (42) give rise to the

steady part of the solution and coordinates ζ_1,ζ_2 of mean particle trajectories obey the steady-state system

$$\frac{d\zeta_1}{dt}=\left\{\frac{\omega\epsilon^2\Pi_f^{-1}}{2F^3(\zeta_2)}\left((2\Pi_\rho-1)\left(\frac{FF'}{2}\sin 2\alpha(\zeta_1,\zeta_2)+\cos 2\alpha(\zeta_1,\zeta_2)\right)+1\right)\times\exp\left(-\frac{\zeta_1^2}{2F^2(\zeta_2)}\right)\right\}\zeta_1,$$
$$\frac{d\zeta_2}{dt}=\frac{\omega\epsilon^2\Pi_f^{-1}}{4F(\zeta_2)}(2\Pi_\rho-1)\sin 2\alpha(\zeta_1,\zeta_2)\times\exp\left(-\frac{\zeta_1^2}{2F^2(\zeta_2)}\right), \tag{44}$$

the right-hand sides of this system may be considered as mean forces, which result from gas oscillations. Equations (44) have the stationary points

$$\zeta_1=0,\qquad \zeta_2=\zeta_2^{(n)}\quad (n=0,1,...), \tag{45}$$

at which the right-hand sides of Eqs. (44) vanish.

The values $\zeta_2^{(n)}$ in Eq. (45) are the roots of the equation

$$\omega_k\zeta_2-\frac{1}{2}\int_0^1\frac{d\tau}{F^{(2)}(\tau)}=\frac{n\pi}{2}\quad (n=0,1,...). \tag{46}$$

An even number n corresponds to a horizontal node line of the vertical fluid velocity component and an add number n corresponds to a horizontal antinode line, Eqs. (32).

Let us linearize system (44) in the neighborhood of stationary points (45). Then the linear system

$$\frac{d\zeta_1}{dt}=\frac{\omega\epsilon^2\Pi_f^{-1}\Pi_\rho}{F^3(\zeta_2^{(n)})}\zeta_1\quad (n=0,2,...),$$
$$\frac{d\zeta_2}{dt}=\frac{\omega\epsilon^2\Pi_f^{-1}}{4F(\zeta_2^{(n)})}(2\Pi_\rho-1)(\zeta_2-\phi_2^{(n)}) \tag{47}$$

describes the mean particle motion near the points of intersection of the horizontal node lines of the vertical fluid velocity component and the axis $\zeta_1=0$.

The linear system

$$\frac{d\zeta_1}{dt}=\frac{\omega\epsilon^2\Pi_f^{-1}}{F^3(\zeta_2^{(n)})}(1-\Pi_\rho)\zeta_1\quad (n=1,3,...)$$
$$\frac{d\zeta_2}{dt}=-\frac{\omega\epsilon^2\Pi_f^{-1}}{4F(\zeta_2^{(n)})}(2\Pi_\rho-1)(\zeta_2-\phi_z^{(n)}) \tag{48}$$

holds near the points of intersection of the horizontal antinode lines of the vertical fluid velocity component and the axis $\zeta_1=0$.

Suppose that

$$\Pi_\rho<\tfrac{1}{2}. \tag{49}$$

The case when Π_ρ is very small corresponds with aerosols. Then the point of intersection of the anti-node lines and the axis $\zeta_1=0$ is an unstable nodal point of Eqs. (48). As time increases, particles drift away from this stationary point. The point of intersection of the node lines and the axis $\zeta_1=0$ is a saddle point of Eqs. (47), where particles approach the node lines and drift along these out of the neighborhood of the axis $\zeta_1=0$. In Fig. 2(a) the qualitative behavior of the particles' trajectories is illustrated schematically. In spite of possible trajectory bends in the region between the node and anti-node lines, particles are repelled from the axis $\zeta_1=0$ after all.

The case when

$$\tfrac{1}{2}<\Pi_\rho<1 \tag{50}$$

can be studied in an analogous way. Now system (47) describes an unstable nodal point and system (48) describes a saddle point. The particle trajectories scheme in Fig. 2(b) are the scheme in Fig. 2(a) shifted $\pi/2$ in the direction of ζ_2. The particles now drift away from the axis $\zeta_1=0$ along fluid velocity anti-node lines.

The third case is

$$\Pi_\rho>1, \tag{51}$$

when the particle density is less than the density of the fluid. Now system (47) describes an unstable nodal point and system (48) describes a stable nodal point. On the scheme in Fig. 2(c) particles drift along the fluid velocity anti-node lines toward the point of intersection of the anti-node lines and the axis $\zeta_1=0$.

IV. CLOSURE

The effect of the ''bouncing ball'' acoustic mode in the motion of dispersed particles has been analyzed. The two-dimensional domain considered is contained between two parallel walls cylindrical in form. Small spherical rigid particles are suspended in the fluid between the walls of the vessel. The main interphase force, exerted by the fluid on the particles, is the Stokes force. Particle concentration was supposed to be small and fluid equations have been treated, neglecting the influence of the particles. The ''bouncing ball'' solution is obtained by an asymptotic analysis, with both the amplitude of the fluid oscillations and the wavelength being small. This solution is used to obtain the steady-state equations of the mean particle motion. It is shown that in addition to a particle drift toward the node or anti-node planes of the fluid velocity, there is a particle drift out of the region where the ''bouncing ball'' mode is concentrated. Particle drift occurs along the node ($\Pi_\rho<0.5$) or anti-node ($\Pi_\rho>0.5$) lines of the fluid velocity. The first case corresponds, for example, to aerosols, when $\Pi_\rho\approx 10^{-3}$; the second case corresponds, for example, to water suspensions, where $\Pi_\rho<1$. When the rigid particles are light ($\Pi_\rho>1$), they collect at the points of intersection of the anti-node lines and the diameter. The characteristic side drift velocity of the aerosol particles is comparable to that of the normal drift. Expressions have been obtained that make it possible to evaluate the normal and side drift velocities of the particles.

This interaction of the two-dimensional sonic mode with dispersed particles may be utilized for the particle separation from fluid and applied in the development of new technologies related to the purification of aerosols and dispersions.

Babich, V. M., and Buldyrev, S. S. (**1991**). *Short-Wavelength Diffraction Theory* (Springer-Verlag, Berlin).
Bogolubov, N. N., and Mitropolski, Yu. A. (**1961**). *Asymptotic Methods in the Theory of Nonlinear Oscillations* (Gordon and Breach, New York).
Buldyrev, V. S. (**1966**). ''Shortwave asymptotic behavior of the eigenfunctions of Helmholtz equation,'' Sov. Phys. Dokl. **10**, 718 [Dokl. Akad. Nauk SSSR **163**, 853 (1961)].
Czyz, H. (**1987**). ''The aerosol particle drift in a standing wave field,'' Arch. Acoust. **12**, 3–4, 199–214.
Czyz, H. (**1990**). ''On the concentration of aerosol particles by means of drift forces in a standing wave field,'' Acoustica **70**, 23–28.
Dain, Y., Vainshtein, P., Fichman, M., and Gutfinger, C. (**1994**). ''Side drift of aerosols in a two-dimensional slowly oscillating sonic field,'' Aerosol. Sci. Technol. **21**, 149–156.
Dain, Y., Fichman, M., Gutfinger, C., Pnueli, D., and Vainshtein, P. (**1995**). ''Dynamics of suspended particles in a two-dimensional high-frequency sonic field,'' J. Aerosol Sci. **26**, 575–594.
Duhin, S. S. (**1960**). ''Theory of the aerosol particle drift in a standing sonic wave,'' Colloid. J. **22**, 128–130.
Fittipaldi, F. (**1979**). ''Partial coagulation by means of ultrasonics,'' Acoustica **41**, 263–266.
Gorkov, L. P. (**1962**). ''On the forces acting on a small particle in an acoustical field in an ideal fluid,'' Sov. Phys. Dokl. **6**, 773 [Dokl. Akad. Nauk SSSR **140**, 88 (1961)].
Hager, F., and Benes, E. (**1991**). ''A summary of all forces acting on spherical particles in a sound field,'' Proceedings of the Ultrasonic International 91 Conference and Exhibition, Le Touquet, France, 1–4 July.
Higashitani, T. (**1981**). ''Migration of suspended particles in plane stationary ultrasonic field,'' Chem. Eng. Sci. **36**, 1877–1882.
King, L. V. (**1934**). ''On the acoustic radiation pressure on spheres,'' Proc. R. Soc. London, Ser. A **147**, 212–240.
Nigmatulin, R. I. (**1990**). *Dynamics of Multiphase Media*. Vols. 1, 2 (Annals of Nuclear Energy, Hemisphere Publishing, London).
Rudnick, J. (**1951**). ''Measurements of the acoustic radiation pressure on a sphere in a standing wave field,'' J. Acoust. Soc. Am. **23**, 633–634.
Trinh, E. H., and Robey, J. L. (**1994**). ''Experimental study of streaming flows associated with ultrasonic levitators,'' Phys. Fluids **6**, 3567–3579.
Vainshtein, P., Fichman, M., and Pnueli, D. (**1992**). ''On the drift of aerosol particles in sonic fields,'' J. Aerosol Sci. **23**, 631–637.
Westervelt, P. J. (**1950**). ''The mean pressure and velocity in a plane acoustic wave in gas,'' J. Acoust. Soc. Am. **22**, 319–327.
Westervelt, P. J. (**1951**). ''The theory of steady forces caused by sound waves,'' J. Acoust. Soc. Am. **23**, 312–315.
Westervelt, P. J. (**1957**). ''Acoustic radiation pressure,'' J. Acoust. Soc. Am. **23**, 26–29.
Yosioka, K., and Kawasima, Y. (**1955**). ''Acoustic radiation pressure on a compressible sphere,'' Acustica **5**, 167–173.

A phase-inverting parabolic concentrator for the generation of negative waves in water

M. T. Carnell, S. J. Barrington, and D. C. Emmony
Physics Department, Loughborough University, Loughborough, Leicestershire LE11 3TU, United Kingdom

(Received 6 October 1996; accepted for publication 3 June 1997)

A method of generating very strong focused negative pressure fields in water for cavitation studies from shock waves produced by an electromagnetic transducer is described. By using a reflecting concentrator with a phase-inverting central element, focal pressures of around −16 MPa from a +5-MPa source have been measured. The combination of an electromagnetic transducer and a phase-inverting parabolic concentrator provides a reliable and reproducible source of cavitation bubbles for study. A qualitative study of the cavitation field produced by the device, using a laser schlieren technique, is presented. © *1997 Acoustical Society of America.*
[S0001-4966(97)00311-1]

PACS numbers: 43.25.Zx, 43.25.Yw, 43.80.Sh [MAB]

INTRODUCTION

Large-amplitude ultrasonic waves have long had applications in medicine in areas such as lithotripsy. However, it has recently been proposed that shock waves may also have a therapeutic effect for some cancers.[1] It is thought that the cytotoxic effect of a shock wave on a tumor cell is caused by the collapse of a micro-bubble near the cell membrane, causing its subsequent rupture. This can inhibit tumor growth and increase the susceptibility of the tumor to preventative drugs. It appears that cytotoxicity is related to the amount of cavitation that occurs in and around the tumor. To increase the level of cavitation in the field, Prat and Arefiev[1] seeded the field with additional micro-bubbles. However, the effectiveness of this technique was limited. The most effective way of generating cavitation is found to be by use of a focused rarefaction wave. The primary negative wave generates large quantities of micro-bubbles while the following positive wave collapses these bubbles. In their work Prat and Arefiev used a modified piezoelectric shock wave generator to produce rarefaction waves directly in water. In this paper an alternative approach of phase inversion of a pressure wave is suggested.

A variety of techniques are used in the focusing of ultrasound.[2] With the exception of self-focusing piezoceramic arrays, the main groups can be classified into either reflection-, refraction-, or interference-type systems. In most cases these devices are used with a shock wave generator, to produce a region of high positive pressure followed by a secondary tensile or rarefaction wave. Problems arise when attempting to modify these devices to produce a high negative pressure field as the primary component of the wavefront. Interference systems only work at a narrow range of frequencies and the use of self-focusing piezoceramic arrays is difficult and costly. As far as the authors are aware, there exists no phase-inversion technique based on a refraction system. Wurster *et al.*[3] have shown that high negative pressures can be obtained using a simple polyurethane foam parabolic reflector. The rarefaction wave is produced by reflection at the pressure release (foam) boundary. This device would probably have little practical application due to the inflexibility in the positioning of the focal region. Other reflection systems that could be modified to produce tensile waves include the twin cone reflector[4,5] and the parabolic concentrator.[4,6]

Using an electromagnetic acoustic transducer (EMAT) and parabolic concentrator in which one element inverts the phase of the pressure wave, it has been possible to produce a highly reliable and reproducible system of generating negative waves. Phase inversion is achieved by reflection at a pressure release boundary while focusing is achieved by the parabolic profile of the boundary and a further reflection from a high-impedance planar surface. The concentrator is characterized by low energy losses, low cost, and the ability to operate over a broad range of frequencies. The large negative pressures generated in the focal region of the concentrator and the flexibility of the position of the focal region make it a good source of mass cavitation for studies in medicine or surface erosion.

I. EXPERIMENTAL

The phase-inverting parabolic concentrator (Fig. 1) is a modification of a design described by Kleesattel.[6] It takes the form of an air-filled parabolic cone surrounded by a planar reflector dish. The cone, with a base diameter of 100 mm and a wall thickness of 2 mm, was produced from 5-mm-thick extruded PMMA sheet using a vacuum thermoforming technique. The PMMA was chosen for its ability to faithfully reproduce and then maintain the desired parabolic profile over its working life. Its fracture resistance, low water absorbtion, and low acoustic impedance were also considered important properties. The far simpler but larger outer reflector was manufactured from mild steel. This ensured both a high reflector efficiency due to the high acoustic impedance of the material relative to that of water and only moderate material and production costs. Its planar reflecting surface has an entrance apperture of 160 mm and is inclined at 25° to the axis of the concentrator. To prevent corrosion a zinc and 8% nickel alloy electroplate coating was also applied to the surface of the reflector.

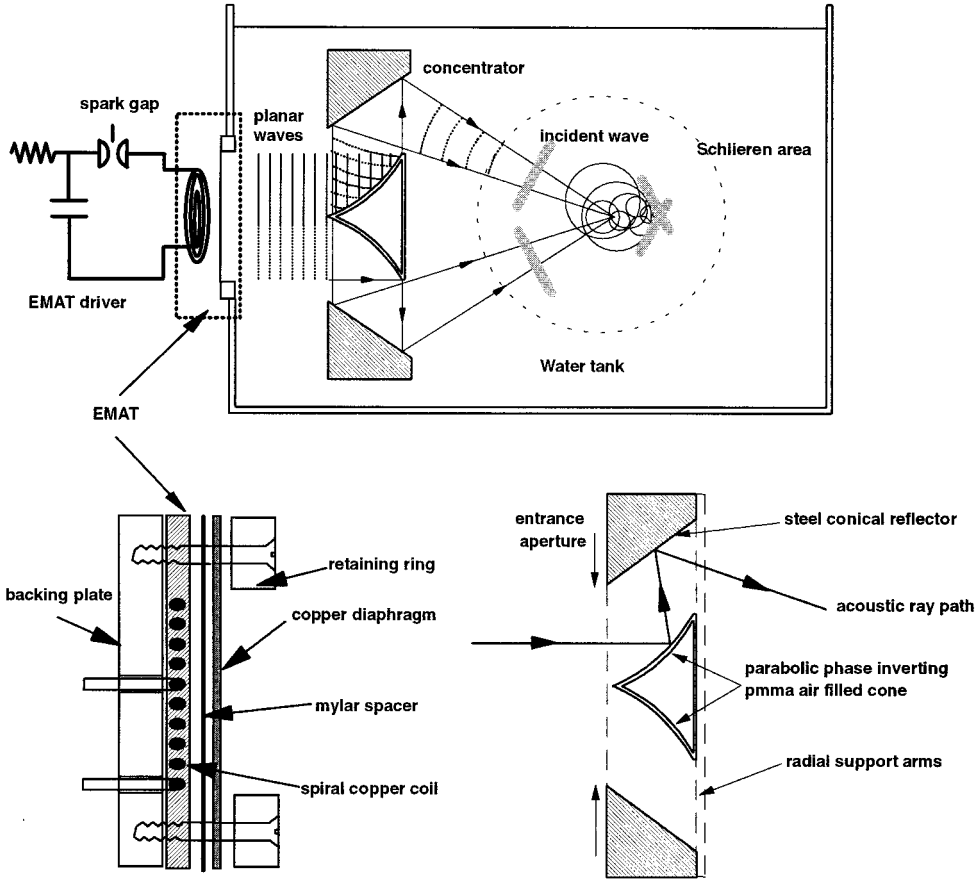


FIG. 1. (a) Layout of the experiment. (b) Cross section of the EMAT. (c) The phase-inverting concentrator.

Planar acoustic waves produced by the EMAT enter the concentrator and undergo reflection and phase inversion at the internal (PMMA/air) interface of the cone (Fig. 2). The converging wavefront then propagates outwards towards the second reflector, where it undergoes a further reflection, before moving towards the focal point 100 mm in front of the concentrator. Initial testing of the concentrator was carried out using an existing acoustic tank and schlieren imaging facility.[7] The tank was filled with laboratory-grade distilled water which was initially partially degassed, but no special precautions were made to prevent adsorption of air at the free surface over the period of experimentation. The two-lens 100-mm-diam parallel beam schlieren system was illuminated by a 20-ns duration Q-switched ruby laser giving a spatial resolution of 30 μm for transients in water. General schlieren observations were carried out using an opaque spot spatial filter, CCD camera, and frame store, since this allowed immediate analysis of the images. For higher resolution images the CCD camera was substituted by a conventional 35-mm camera loaded with Ilford FP4 plus roll film.

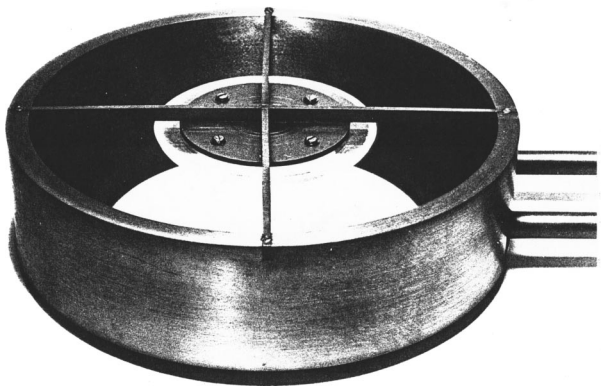


FIG. 2. The parabolic concentrator.

A hydrophone study of the acoustic field of the concentrator was undertaken using a similar technique to that used previously.[8] The impulse response hydrophone was a 3-mm-diam 110-μm-thick PVDF piezopolymer mounted on the end of an unpoled PVDF rod of the same diameter. The theoretical response of this transducer may be estimated from its thickness divided by the velocity of sound in the material, leading to a rise time of <20 ns and a bandwidth of ~ 50 MHz. Absolute calibration of these devices has been obtained previously using interferometrically derived shock wave pressures around individual cavitation transients[9] and by cross calibration against a standard transducer calibrated at the National Physical Laboratory. The agreement between these two techniques allows us to place an uncertainty of less than 10% on our pressure measurements.

II. RESULTS

Previous hydrophone measurements of the acoustic pulses generated by the EMAT[8] have shown it to be both a reliable and reproducible source of shock waves. These ear-

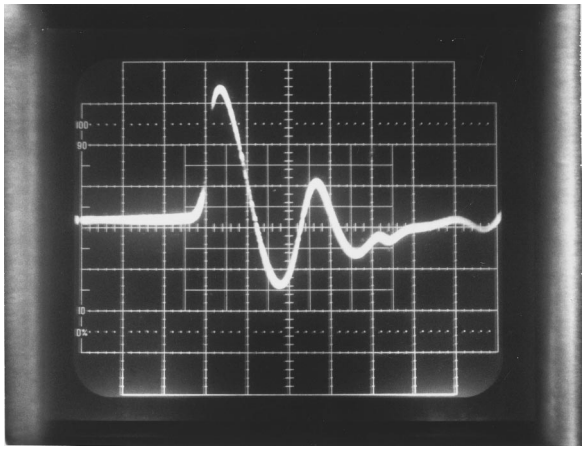


FIG. 3. Transducer trace of the planar EMAT shock wave (1 horizontal div=2 μm, 1 vertical div=2 MPa).

lier measurements have indicated that the peak pressure and period of the planar waves produced by the EMAT (Fig. 3) are in the region of 5 MPa and 6 μs, respectively. The addition of a focusing biconcave lens was found to increase the pressure in the focal zone to around 20 MPa.

The latest hydrophone studies of the focal region produced by the parabolic concentrator indicate that both phase inversion and focusing has taken place. The hydrophone trace in Fig. 4 shows one such pulse at or about the focal region of the concentrator. The relatively short duration negative pulse (approximately 2 μs long) is followed by a longer duration (4 μs) positive transient, the leading edge of which is beginning to exhibit a fast rising, shocklike, characteristic. The peak negative pressure is estimated to be around -16 MPa and the peak positive pressure in the second part around 10 MPa. The total period of the wave is, as expected, unchanged at 6 μs. By taking a number of individual pressure measurements at fixed intervals around the field it has been possible to build up a map of the focal region. In the 2-D contour map of Fig. 5, the focal zone is approximately 10 mm$\times$50 mm. This is remarkably close to the focal dimensions achieved using the earlier biconcave acoustic lens[7] and has been produced despite the concentrator only having half the effective aperture of the lens.

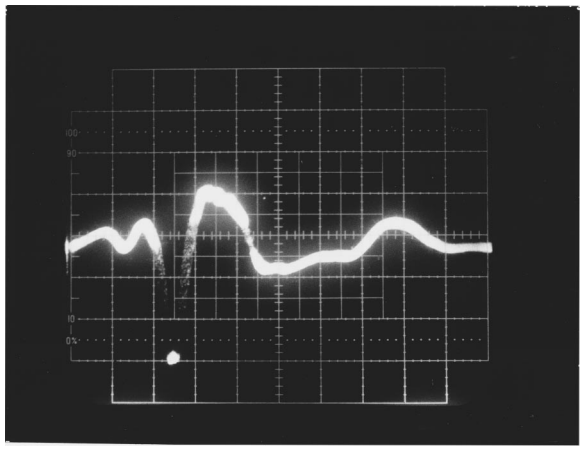


FIG. 4. Transducer trace of the focused phase-inverted wavefront at the focal point of the concentrator (1 horizontal div=2 μs, 1 vertical div =5 MPa).

The results of the schlieren study for a negative pressure wave in the focal region of the parabolic concentrator are shown in Fig. 6(a)–(d). In the first of these images [Fig. 6(a)] we see the wavefront a short time after emerging from the concentrator. The saucer-shaped wavefront is observed edge on and appears in the image as two bright but distinct components corresponding to the upper and lower sections of the concentrator. Light passing through the central part of the field only encounters a thin shell of the wavefront (in front and behind the plane of the photograph) and therefore passes though the working section largely undisturbed. Light taking this path will ultimately be removed by the schlieren system filter and be lost to the system. Only light passing through a part of the wavefront tangentially, such as that at the top and bottom, will be imaged. At this point in time no cavitation in the field is visible. In the next image of this sequence [Fig. 6(b)] we see the tips of the two components come together in the center of the field. On the left are the circular structures associated with cavitation transients. These are brought about by the collapse of micro-bubbles generated in the water by the wave as it passes through the region. As expected, the large tensile wave has resulted in the production of a large amount of cavitation activity. The level of cavitation activity is further increased when the wave front reaches the focal point [Fig. 6(c)] approximately 100 mm in front of the concentrator. The cavitation transients expand symmetrically with time, so that in general the largest diameter transients, such as those on the left of the image, have been produced by earlier cavitation collapse events. The overall appearance of the cavitation transients is thus one of a fan like structure made up of many cavitation transients travelling behind the wavefront. However, some smaller cavitation transients are also visible within parts of the field dominated by the earlier events, particularly in Fig. 6(b). It is suggested that these are secondary cavitation events,[10] perhaps rebounding bubbles, that have been caused by the interaction between the initial negative wave and the following shock wave with microbubbles already present in the field. This very same phenomenon has been observed in schlieren studies of the field of a conventional EMAT shock wave.

In the final image of the sequence, Fig. 6(d), the wavefront is observed having passed beyond the focal point. The two negative wave components have crossed in the field and are now expanding. In the image the results of extensive cavitation are clearly evident by the amount of scattered light produced by the many transients. Throughout the central region and especially visible within the wavefronts of Fig. 6(c) and (d) are black spots. In earlier investigations[10] it has been suggested that these are also cavitation bubbles, although with estimated diameters of between 100–400 μm the exact growth mechanism is uncertain.

The study into the production and imaging of the tensile waves obtained in the focal region of our phase-inverting parabolic concentrator has relied on the use of an existing EMAT source. The operational specifications of this EMAT, particularly the relatively low discharge energy of 81 J, has to a certain extent limited the ultimate pressures obtained in the focal region of the concentrator. It is expected that a

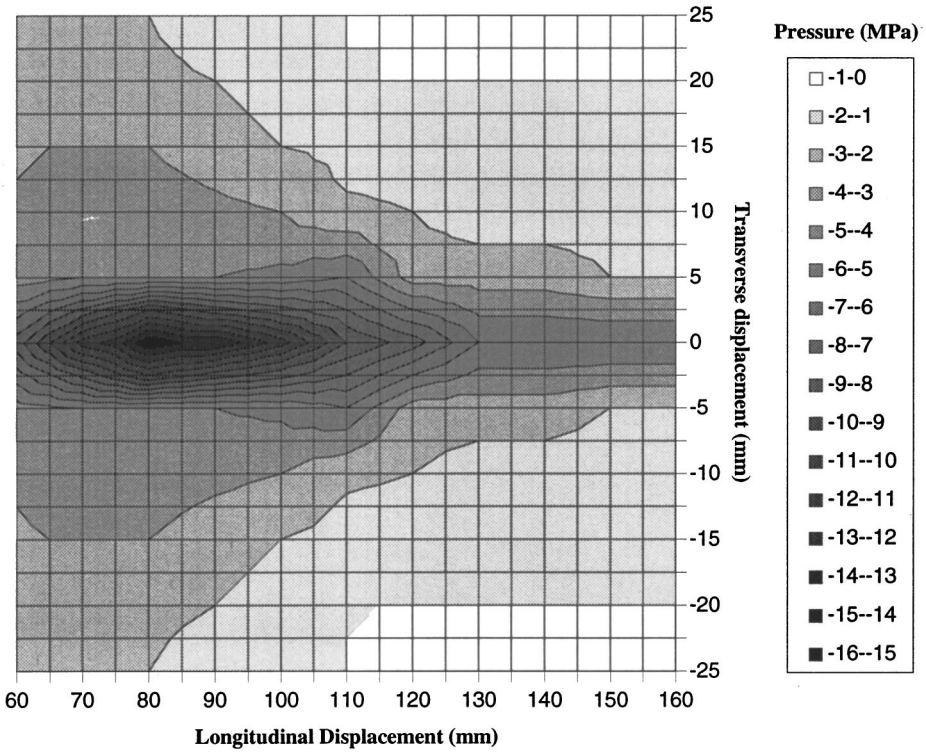


FIG. 5. Pressure map of the focal region.

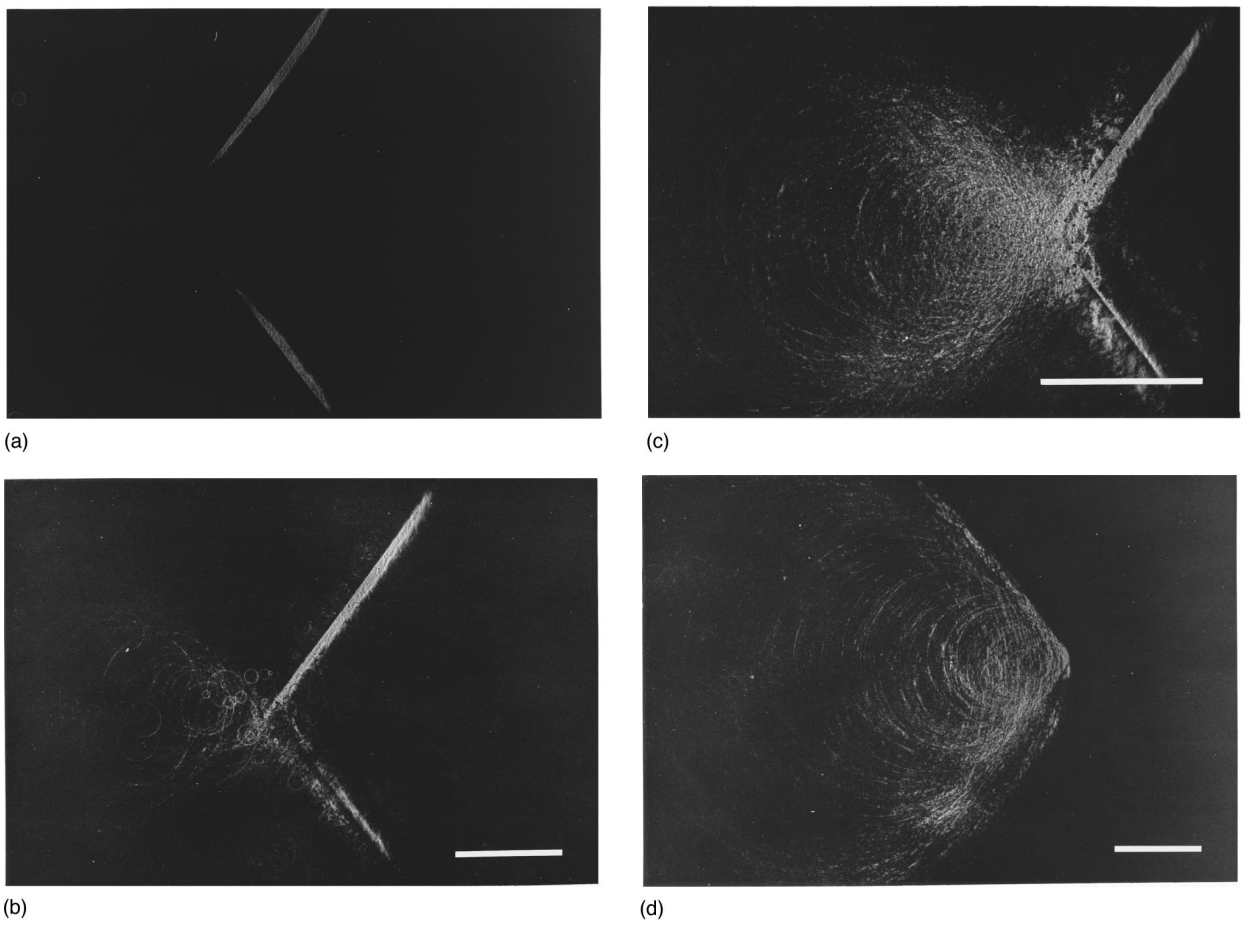


FIG. 6. Schlieren images of the focused phase-inverted wavefront, propagation is from left to right. (a) Phase-inverted wavefront emerging from the concentrator. (b) The phase-inverted wavefront approaching the focal region. (c) The phase-inverted wavefront at the focal point of the system. (d) The phase-inverted wavefront beyond the focal point. The scale bar in each image represents 10 mm.

significant increase in the magnitude of the tensile wave would be obtained if a higher-energy/commercial EMAT were used. So far the damage potential of this cavitation field has still to be investigated.

III. CONCLUSION

The schlieren study of the focused negative waves shows that the EMAT and the phase-inverting parabolic concentrator provides a reliable and reproducable source of high-amplitude tensile waves in water. The compact design and cost effectiveness set it apart from the alternative methods of generating negative pressures, such as those described by Prat and Arefiev[1] and Wurster *et al.*[3] The results of the schlieren studies have shown that the combination of a strong negative pulse followed by a positive component with a shocklike leading edge increases the level of cavitation observed in the field compared to a conventional EMAT pulse,[10] where the compressive pulse is followed by a negative pressure. With a usable focal region this makes it an ideal source for cavitation studies. Possible applications include its use to examine the effects of cavitation on tissue and stones during lithotripsy and for investigations into the cytotoxic effects of cavitation on tumors in the treatment of cancer. The device is also a possible source for cavitation erosion studies.

ACKNOWLEDGMENTS

This work was supported by a start up grant from Loughborough University. The authors gratefully acknowledge the assistance of Professor Birley, Andrew Trotter, and Trevor Downham of Loughborough University for their help in the manufacture of the concentrator.

[1] F. Prat and A. Arefiev, ''Cellular and tissular effects of shock wave induced cavitation: Potential application to digestive cancers,'' Proceedings Shock Waves, Marseille, France (1995), pp. 21–28.

[2] T. Tarnoczy, ''Sound focusing lenses and waveguides,'' Ultrasonics **3**, 115–127 (1965).

[3] C. Wurster *et al.*, ''Negative pressure measurements of water using the glass fibre optic hydrophone,'' World Congress on Ultrasonics (1995), pp. 635–638.

[4] A. Barone, ''Aspects of the concentration of ultrasonic energy,'' Acustica **2**, 221–225 (1952).

[5] J. Pahud, ''Etude experimentale du champ de transducteurs ultrasonores focalisants,'' Acustica **4**, 205–207 (1954).

[6] C. Kleesattel, ''Bemerkung zu der Abhandlung: A Barone, Aspects of the concentration of ultrasonic energy,'' Acustica **3**, 407 (1953).

[7] M. T. Carnell, R. D. Alcock, and D. C. Emmony, ''Optical imaging of shock waves produced by a high energy electromagnetic transducer,'' Phys. Med. Biol. **38**, 1575–1588 (1993).

[8] M. T. Carnell and D. C. Emmony, ''Optical distortion in the field of a lithotripter shock wave,'' J. Appl. Opt. **34**, 6465–6470 (1995).

[9] B. Ward and D. C. Emmony, ''Interferometric studies of the pressures developed in a liquid during infrared laser-induced cavitation-bubble oscillation,'' Infrared Phys. **32**, 489–515 (1991).

[10] M. T. Carnell, P. T. Fiadeiro, and D. C. Emmony, ''Cavitation phenomena generated by a lithotripter shock wave,'' J. Acoust. Soc. Am. **97**, 677–679 (1995).

Propagation of sound in a turbulent medium. I. Plane waves

V. E. Ostashev[a)]
Department of Physics, New Mexico State University, Las Cruces, New Mexico 88003-8001

Volker Mellert, Ralf Wandelt, and Frank Gerdes
Arbeitsgruppe Akustik, Fachbereich Physik, Universität Oldenburg, Oldenburg D-26111, Germany

(Received 20 December 1995; revised 9 June 1997; accepted for publication 7 July 1997)

The paper presents formulas for the statistical moments of a plane sound wave propagating in moving random media (the turbulent atmosphere and ocean, turbulent flows of gases and fluids, etc.) with arbitrary spectra of temperature and medium velocity fluctuations. These statistical moments, which are most often of interest, are the following: the variances of log-amplitude and phase fluctuations, the correlation functions of log-amplitude and phase fluctuations, the mean sound field, the coherence function of a sound field, and the sound scattering cross section per unit volume. These statistical moments are calculated for Gaussian spectra of temperature and medium velocity fluctuations. It is shown that the contributions to the statistical moments due to sound scattering by medium velocity fluctuations may dramatically differ from those due to sound scattering by temperature fluctuations. © *1997 Acoustical Society of America.* [S0001-4966(97)04710-3]

PACS numbers: 43.28.Fp, 43.20.Bi, 43.20.Fn [LCS]

INTRODUCTION

Three assumptions have often been used when calculating the statistical moments of a sound field propagating in moving random media (the turbulent atmosphere and ocean, turbulent flows of gases and fluids, etc.). The first assumption was that $L_0 > \lambda$, where L_0 is the scale of largest inhomogeneities in a medium, which affect these statistical moments, and λ is the sound wavelength. In this case and for line-of-sight sound propagation, the sound pressure $p(x,\mathbf{r})$ can be sought as a solution of the parabolic equation[1,2]

$$2ik\,\frac{\partial p}{\partial x}+\Delta_\perp p+2k^2\left(1+\frac{\epsilon_{\rm mov}}{2}\right)p=0. \tag{1}$$

Here, a sound wave propagates along the x axis, $\mathbf{r}=(y,z)$ are the transverse coordinates, $k=2\pi/\lambda$ is the wave number, $\Delta_\perp=(\partial^2/\partial y^2,\partial^2/\partial z^2)$, and $\epsilon_{\rm mov}=\epsilon-2v_x/c_0$ are fluctuations in the square of the acoustic refractive index in a moving random medium. In the expression for $\epsilon_{\rm mov}$, v_x are fluctuations of the medium velocity component in the direction of sound propagation, $\epsilon=c_0^2/(c_0+\tilde{c})^2-1=-\beta_c\tilde{T}/T_0$ is the random deviation of the square of the refractive index in a motionless random medium from 1, where c_0 and $\tilde{c}$ are the mean value and fluctuating part of the adiabatic sound speed c, T_0 and $\tilde{T}$ are the mean value and fluctuating part of the temperature T, and the coefficient $\beta_c=2(T_0/c_0)\times(\partial c_0/\partial T_0)$. In the atmosphere $\beta_c=1$, while in the ocean β_c depends on T_0 and is greater than 1.[2] For homogeneous and isotropic turbulence, the random fields ϵ and v_x do not correlate.[3,4] Since $\epsilon=-\beta_c\tilde{T}/T_0$, statistical characteristics of ϵ coincide with those for temperature fluctuations $\tilde{T}$. Therefore, when considering the random field ϵ, we shall often relate it to temperature fluctuations.

The second assumption was to utilize the Kolmogorov or Gaussian spectrum for describing temperature and medium velocity fluctuations (i.e., the random fields ϵ and v_x). The Kolmogorov spectrum adequately describes small-scale inhomogeneities in the inertial range of locally homogeneous and isotropic turbulence. In the 1950–1970's, this spectrum was widely used to calculate the statistical moments of a sound field propagating through the turbulent atmosphere.[4,5] The Gaussian spectrum allows one to approximately take into account the effects of largest inhomogeneities in a medium on sound propagation[6] and is very convenient for analytical calculations. Since the end of 1970's and prior to the beginning of 1990's, this spectrum was almost exclusively used for calculating the statistical moments of a sound field in the turbulent atmosphere.[7–13] Furthermore, the Gaussian spectrum has also been used in oceanic acoustics.[14,15] Finally, this spectrum approximates adequately large-scale inhomogeneities in turbulent jets and over heated grids.

Note that an anisotropic Gaussian spectrum has been used in Ref. 16 to describe the source region of atmospheric turbulence. Also it has been shown in Ref. 16 that the Gaussian spectrum of atmospheric turbulence employed in Refs. 8, 9, and 12 cannot be used to represent the source region. Nevertheless, the latter spectrum can probably be used for calculating certain statistical moments of a sound field if these moments are affected by inhomogeneities which are much less than those in the source region. This can be the case if a time interval for averaging sound field fluctuations is relatively small, or a statistical moment of a sound field ''feels'' only inhomogeneities which are much less than those in the source region because of the geometry of a problem.

Both aforementioned assumptions are valid for many problems of sound propagation in random media. But for the third assumption, it was actually assumed in the previous theories of sound propagation in the turbulent atmosphere (e.g., Refs. 4, 5, 7–12) that v_x can be considered as a scalar random field with similar statistical properties as those for the scalar random field ϵ. However, it has been shown in

[a)]On leave from Institute of Atmospheric Physics, Moscow 109017, Russia.

Refs. 1, 2, and 17 that this is not the case. Indeed, v_x is a component of a vector random field $\mathbf{v}$, where $\mathbf{v}=(v_x,v_y,v_z)$ is the medium velocity vector. Therefore, statistical properties of v_x are different from those for the scalar random field ϵ. For example, for homogeneous and isotropic turbulence, v_x must have an anisotropic correlation function, while the correlation function of ϵ is isotropic.

For the Kolmogorov spectrum, the third assumption leads to the result that the so-called effective structure parameter $C_{\rm eff}^2$ is given by $C_{\rm eff}^2=C_\epsilon^2+4C_v^2/c_0^2$, where C_ϵ^2 and C_v^2 are the structure parameters of ϵ and $\mathbf{v}$. $C_{\rm eff}^2$ appears in all equations for the statistical moments of a sound field. The formula for $C_{\rm eff}^2$, given above, has been adopted in the previous theories.[4,5] However, taking into account that v_x is a component of a vector random field, it has been shown[2,17] that the correct formula for $C_{\rm eff}^2$ is given by $C_{\rm eff}^2=C_\epsilon^2+(22/3)C_v^2/c_0^2$. This formula differs from that adopted in the previous theories by a numerical factor multiplying C_v^2. This factor is 22/3 in the correct formula for $C_{\rm eff}^2$ and is 4 in that used in the previous theories. The factor 22/3 of $C_{\rm eff}^2$ has been verified experimentally for the case of sound propagation in the turbulent ocean.[18]

The equations for the statistical moments of the sound field p in a motionless random medium with a Gaussian correlation function of ϵ are well-established in the literature.[4,6,14,15] All of these equations contain the variance σ_ϵ^2 of the random field ϵ. The third assumption and the fact that in Eq. (1) $\epsilon_{\rm mov}=\epsilon-2v_x/c_0$ immediately lead to the result that the equations for the statistical moments of p in a moving random medium with a Gaussian correlation function for medium inhomogeneities coincide with the equations for the analogous statistical moments of p in a motionless random medium, if in the latter equations σ_ϵ^2 is replaced by $\sigma_{\rm mov}^2=\sigma_\epsilon^2+4\sigma_v^2/c_0^2$. Here, $\sigma_{\rm mov}^2$ and σ_v^2 are the variances of the random fields $\epsilon_{\rm mov}$ and v_x, respectively. This result has also been adopted in the previous theories. However, taking into account that v_x is a component of a vector random field, it has been shown[2] that the equations for the statistical moments of p in a moving random medium may not be obtained by the replacement of σ_ϵ^2 by $\sigma_{\rm mov}^2$ and should be calculated anew. In Ref. 2, only the variances of log-amplitude and phase fluctuations, $\langle\chi^2\rangle$ and $\langle\phi^2\rangle$, of a plane wave propagating in a moving random medium with the Gaussian correlation function of medium inhomogeneities have been calculated. Here, χ and ϕ are fluctuations in log-amplitude and phase of a sound wave, which are related to p by $p=p_0\exp(ikx+\chi+i\phi)$, where p_0 is a solution of Eq. (1) for the case $\epsilon_{\rm mov}=0$. Note that $p_0=\exp(ikx)$ for plane wave propagation, and $p_0=\exp(ikR)/4\pi R$ for spherical wave propagation, where R is the modulus of the vector $\mathbf{R}=(x,\mathbf{r})$.

The main goal of the present paper and the subsequent one,[19] which is referred to hereinafter as Paper II, is to calculate other statistical moments of the sound field p propagating in a moving random medium, namely, the correlation functions of log-amplitude and phase fluctuations, $B_\chi(\mathbf{r})=\langle\chi(x,\mathbf{r}_1)\chi(x,\mathbf{r}_1+\mathbf{r})\rangle$ and $B_\phi(\mathbf{r})=\langle\phi(x,\mathbf{r}_1)\phi(x,\mathbf{r}_1+\mathbf{r})\rangle$, the mean sound field $\langle p(x,\mathbf{r})\rangle$, the transverse coherence function $\Gamma(\mathbf{r}_1,\mathbf{r}_2)=\langle p(x,\mathbf{r}_1)p^*(x,\mathbf{r}_2)\rangle$, and the sound scattering cross section per unit volume σ. Moreover, on the basis of the equations obtained, the temperature and medium velocity contributions to these statistical moments are compared. In this paper we shall calculate the statistical moments of a plane wave. The statistical moments of a spherical wave will be calculated in Paper II. Note that the study of plane wave propagation is as a rule separated from that for a spherical wave in the classical treatments of waves in random media (e.g., see Refs. 4, 6, 14, 15).

In Secs. I A–I G, we shall consider line-of-sight sound propagation in a moving random medium, where $L_0>\lambda$. In Sec. I B, formulas for the statistical moments of p in a medium with arbitrary (anisotropic) spectra of temperature and medium velocity fluctuations are presented. In Secs. I C–I G, these statistical moments are calculated and analyzed in detail for the Gaussian spectra. The attenuation coefficient of the mean sound field in a moving random medium with $L_0<\lambda$ is calculated in Sec. I H. The sound scattering cross section is obtained in Sec. II. And in Sec. III, the obtained results are summarized and discussed.

I. LINE-OF-SIGHT SOUND PROPAGATION

A. Preliminary remarks

For line-of-sight sound propagation, the statistical moments of p, which are usually dealt with, are the following: $\langle\chi^2\rangle$, $\langle\phi^2\rangle$, B_χ, B_ϕ, $\langle p\rangle$, and Γ. If $v_x=0$, the formulas for these statistical moments have been obtained in the literature in the Markov approximation for the case of homogeneous anisotropic turbulence. All of these formulas contain $\Phi_\epsilon(0,\mathbf{K}_r)$, where $\Phi_\epsilon(K_x,\mathbf{K}_r)$ is the three-dimensional spectral density of the random field $\epsilon(x,\mathbf{r})$, which is related to the correlation function $B_\epsilon(x,\mathbf{r})$ of ϵ by

$$B_\epsilon(x,\mathbf{r})=\langle\epsilon(x_1,\mathbf{r}_1)\epsilon(x_1+x,\mathbf{r}_1+\mathbf{r})\rangle$$
$$=\int dK_x\int d^2K_r\, e^{iK_xx+i\mathbf{K}_r\cdot\mathbf{r}}\Phi_\epsilon(K_x,\mathbf{K}_r), \qquad (2)$$

where K_x and $\mathbf{K}_r$ are the components of the wave number vector. It has been shown[2] that, if in these formulas for the statistical moments of the sound field $\Phi_\epsilon(0,\mathbf{K}_r)$ is replaced by

$$\Phi_{\rm eff}(0,\mathbf{K}_r)=\Phi_\epsilon(0,\mathbf{K}_r)+(4/c_0^2)\Phi_{xx}(0,\mathbf{K}_r), \qquad (3)$$

one obtains formulas for the corresponding statistical moments in a moving random medium. Here, $\Phi_{xx}(K_x,\mathbf{K}_r)$ is a three-dimensional spectral density of $v_x(x,\mathbf{r})$ and is related to the correlation function $B_{xx}(x,\mathbf{r})$ of v_x by a formula analogous to Eq. (2). Such an approach allows one to obtain formulas for $\langle\chi^2\rangle$, $\langle\phi^2\rangle$, B_χ, B_ϕ, $\langle p\rangle$, and Γ for the case of plane wave propagation in homogeneous anisotropic turbulence with arbitrary spectra of temperature and medium velocity fluctuations. Since these formulas are important in theories of sound propagation in turbulent media, they are presented in the next subsection.

B. General formulas

The formulas for the correlation functions of log-amplitude and phase fluctuations of a sound wave in a mov-

ing random medium, calculated by the Rytov method, can be obtained from formulas for the corresponding correlation functions in a motionless random medium[6] and by using Eq. (3)

$$B_{\chi,\phi}(r)=\frac{\pi^2k^2x}{2}\int_0^\infty\left(1\mp\frac{K_F^2}{K^2}\sin\frac{K^2}{K_F^2}\right)\times J_0(Kr)\Phi_{\text{eff}}(0,K)KdK. \tag{4}$$

Here, x is the distance of sound propagation in a random medium, J_0 is the Bessel function, K is a modulus of $\mathbf{K}_r$, and $K_F^2=k/x$. The upper and lower signs on the right-hand side of Eq. (4) correspond to B_χ and B_ϕ, respectively. The formula for B_χ is valid only in the weak scatter limit when $\langle\chi^2\rangle\ll 1$.

Since $\langle\chi^2\rangle=B_\chi(r=0)$ and $\langle\phi^2\rangle=B_\phi(r=0)$, the variances of log-amplitude and phase fluctuations are given by Eq. (4) with $r=0$.

The expression for the mean sound field in a moving random medium, calculated by the parabolic equation method, can be obtained from the corresponding expression in a motionless random medium[6] and Eq. (3):

$$\langle p(x,\mathbf{r})\rangle=\exp(-\gamma x)p_0(x,\mathbf{r}). \tag{5}$$

Here, $p_0=\exp(ikx)$ for the considered case of plane wave propagation, and γ is the attenuation coefficient of the mean field due to sound scattering by medium inhomogeneities, given by

$$\gamma=\frac{\pi^2k^2}{2}\int_0^\infty\Phi_{\text{eff}}(0,K)KdK. \tag{6}$$

It follows from Eq. (5) that $\langle p\rangle$ decreases exponentially if x is increased. The decrease in $\langle p\rangle$ is accompanied by simultaneous increase in the fluctuating part of the sound field $\tilde{p}=p-\langle p\rangle$ so that the intensity of the sound field remains constant.[6]

The coherence function $\Gamma(\mathbf{r}_1,\mathbf{r}_2)$ of a plane wave depends only on the modulus of the vector $\mathbf{r}=\mathbf{r}_1-\mathbf{r}_2$. The formula for $\Gamma(r)$ in a moving random medium, calculated by the parabolic equation method, can be obtained from the corresponding formula in a motionless random medium[6] and Eq. (3):

$$\Gamma(r)=I_0\exp\left\{-\pi^2k^2x\int_0^\infty(1-J_0(\kappa r))\Phi_{\text{eff}}(0,K)KdK\right\}, \tag{7}$$

where $I_0=p_0p_0^*$ is proportional to the sound intensity in a medium without random inhomogeneities.

C. Effective spectral density

Equations (4)–(7) are valid for arbitrary (anisotropic) spectra of medium inhomogeneities. For the Gaussian spectra of temperature and medium velocity fluctuations, the effective spectral density Φ_{eff} in these equations is given by[2]

$$\Phi_{\text{eff}}(0,K)=[\sigma_\epsilon^2+(K^2l^2/4)(4\sigma_v^2/c_0^2)](l^3/8\pi^{3/2})\times\exp(-K^2l^2/4). \tag{8}$$

Here, l is the correlation length of medium inhomogeneities. In the square brackets in Eq. (8), the factor $K^2l^2/4$ appears before $4\sigma_v^2/c_0^2$ due to the fact that v_x is a component of the solenoidal vector random field $\mathbf{v}$. This factor distinguishes the temperature and medium velocity contributions to Φ_{eff}. In the previous theories, this factor was actually assumed to be equal to 1.

D. Correlation functions for log-amplitude and phase fluctuations

We substitute $\Phi_{\text{eff}}(0,K)$ from Eq. (8) into Eq. (4), calculate the integrals over K in the equations obtained (the calculation of these integrals is rather complicated and is presented in the Appendix), and express B_χ and B_ϕ as the sums: $B_\chi=B_\chi^\epsilon+B_\chi^v$ and $B_\phi=B_\phi^\epsilon+B_\phi^v$. Here B_χ^ϵ and B_χ^v describe the temperature and medium velocity contributions to B_χ, and B_ϕ^ϵ and B_ϕ^v describe the temperature and medium velocity contributions to B_ϕ. As a result, we get the formulas for the correlation functions for log-amplitude and phase fluctuations

$$B_\chi(r)=B_\chi^\epsilon+B_\chi^v=\frac{\sqrt{\pi}k^2lx}{8}\left\{\left[e^{-r^2/l^2}+\frac{1}{D}\,\Im E_1\left(\frac{r^2e^{i\arctan D}}{l^2\sqrt{1+D^2}}\right)\right]\sigma_\epsilon^2+\left[\left(1-\frac{r^2}{l^2}\right)e^{-r^2/l^2}-\frac{\exp\left(-\frac{r^2}{l^2(1+D^2)}\right)}{1+D^2}\left[\cos\left(\frac{r^2D}{l^2(1+D^2)}\right)-\frac{1}{D}\sin\left(\frac{r^2D}{l^2(1+D^2)}\right)\right]\right]\frac{4\sigma_v^2}{c_0^2}\right\}, \tag{9}$$

$$B_\phi(r)=B_\phi^\epsilon+B_\phi^v=\frac{\sqrt{\pi}k^2lx}{8}\left\{\left[e^{-r^2/l^2}-\frac{1}{D}\,\Im E_1\left(\frac{r^2e^{i\arctan D}}{l^2\sqrt{1+D^2}}\right)\right]\sigma_\epsilon^2+\left[\left(1-\frac{r^2}{l^2}\right)e^{-r^2/l^2}+\frac{\exp\left(-\frac{r^2}{l^2(1+D^2)}\right)}{1+D^2}\left[\cos\left(\frac{r^2D}{l^2(1+D^2)}\right)-\frac{1}{D}\sin\left(\frac{r^2D}{l^2(1+D^2)}\right)\right]\right]\frac{4\sigma_v^2}{c_0^2}\right\}. \tag{10}$$

Here, $D=(2/K_Fl)^2=4x/(kl^2)$ is the wave parameter widely used in theories of waves in random media,[6] and $E_1(x)=\int_x^\infty y^{-1}e^{-y}dy$ is the exponential-integral function, the imaginary part of which can be expressed in the form

$$\Im E_1\left(\frac{r^2 e^{i\ \arctan D}}{l^2\sqrt{1+D^2}}\right)=-\int_{r^2/(l^2(1+D^2))}^{\infty} y^{-1}e^{-y}\ \sin(yD)dy. \tag{11}$$

The formulas for B_χ^ϵ and B_ϕ^ϵ are also given in Ref. 14 where another approach is used for their derivation. [In Ref. 14 these formulas are written in a form which is different from that in Eqs. (9) and (10); however, calculations show that both forms are equivalent.] The formulas for B_χ^v and B_ϕ^v are new results obtained in the present paper.

When comparing the temperature and medium velocity contributions to the statistical moments of p, we shall always assume that

$$\sigma_\epsilon^2=4\sigma_v^2/c_0^2. \tag{12}$$

In this case, temperature and medium velocity fluctuations make the same contributions to the variance $\sigma_{\rm mov}^2=\sigma_\epsilon^2+4\sigma_v^2/c_0^2$ of the random field $\epsilon_{\rm mov}$ which appears in Eq. (1). It is clearly seen from Eqs. (9) and (10) that B_χ^ϵ significantly differs from B_χ^v, and B_ϕ^ϵ significantly differs from B_ϕ^v. In other words, the temperature and medium velocity contributions to the correlation functions B_χ and B_ϕ are different, and these functions are not proportional to $\sigma_{\rm mov}^2=\sigma_\epsilon^2+4\sigma_v^2/c_0^2$ as it was previously thought.

It is worthwhile obtaining the two asymptotes of Eqs. (9) and (10). To get the first asymptotic, we shall assume that

$$D\ll 1,\quad Dr^2/l^2\ll 1. \tag{13}$$

Note that a value of r much greater than l is not of any practical interest since $B_\chi(r)$ and $B_\phi(r)$ are nearly zero for $r\gg l$ (see Figs. 1 and 2). On the other hand, if $r<l$, the two inequalities in Eq. (13) can be replaced by one, $D\ll 1$. The latter inequality is for a range of applicability of the geometric-acoustic approximation.[6] In Eqs. (9)–(11), we expand cosines, sines, exponents, and other functions into Taylor series with respect to the small parameters D and Dr^2/l^2. Retaining only the lowest order nonvanishing terms yields the asymptotics of the correlation functions

$$B_\chi(r)=B_\chi^\epsilon+B_\chi^v=\frac{2\sqrt{\pi}x^3}{3l^3}\left[\left(1-\frac{2r^2}{l^2}+\frac{r^4}{2l^4}\right)\sigma_\epsilon^2+\left(1-\frac{3r^2}{l^2}+\frac{3r^4}{2l^4}-\frac{r^6}{6l^6}\right)\frac{12\sigma_v^2}{c_0^2}\right]e^{-r^2/l^2}, \tag{14}$$

$$B_\phi(r)=B_\phi^\epsilon+B_\phi^v=\frac{\sqrt{\pi}k^2lx}{4}\left[\sigma_\epsilon^2+\left(1-\frac{r^2}{l^2}\right)\frac{4\sigma_v^2}{c_0^2}\right]e^{-r^2/l^2}. \tag{15}$$

To obtain the second asymptotic, we shall assume that

$$D\gg 1,\quad Dl^2/r^2\gg 1. \tag{16}$$

Since a value of r, of the order of or smaller than l, is usually of interest, the two inequalities in Eq. (16) can be replaced by one $D\gg 1$. The latter inequality is for a range of applicability of the Fraunhofer diffraction. Expanding all functions in Eqs. (9)–(11) into Taylor series in the small parameters $1/D$ and r^2/l^2D^2 and retaining only the lowest order nonvanishing terms yields the following asymptotic approximations of the correlation functions:

$$B_\chi(r)=B_\phi(r)=\frac{\sqrt{\pi}k^2lx}{8}\left[\sigma_\epsilon^2+\left(1-\frac{r^2}{l^2}\right)\frac{4\sigma_v^2}{c_0^2}\right]e^{-r^2/l^2}. \tag{17}$$

Equations (14), (15), and (17) give us relatively simple formulas for B_χ and B_ϕ for two important limiting cases. Another limiting case is $r=0$. In this case, Eqs. (9) and (10) give us the variances for log-amplitude and phase fluctuations

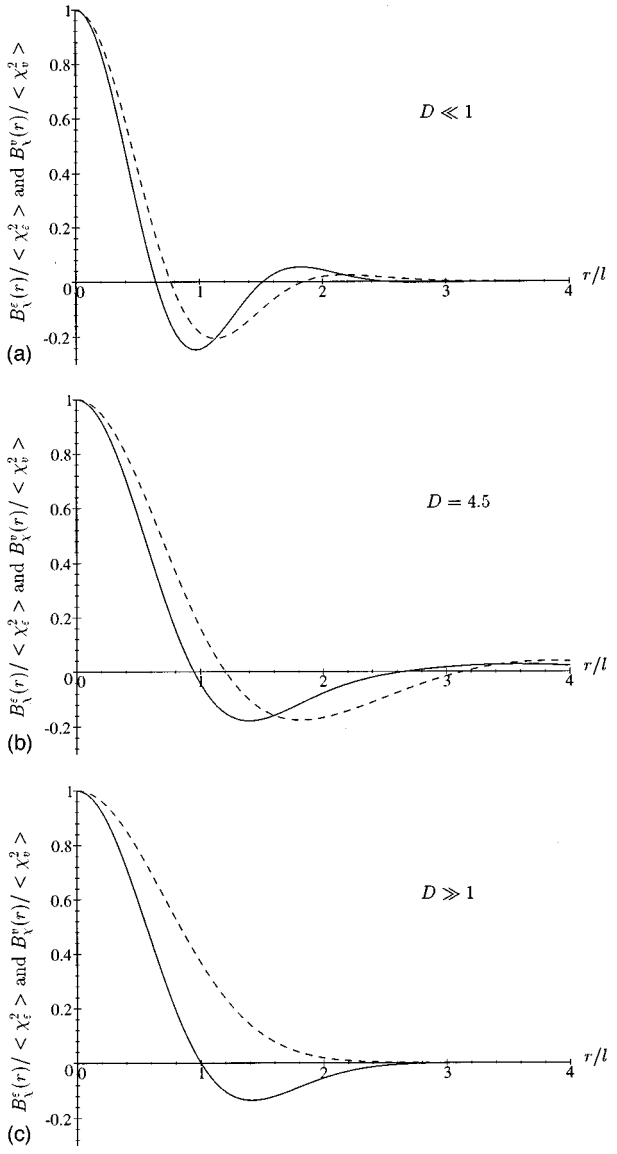


FIG. 1. The normalized temperature and medium velocity contributions to the correlation function for log-amplitude fluctuations, $B_\chi^\epsilon(r)/\langle\chi_\epsilon^2\rangle$ (dashed line) and $B_\chi^v(r)/\langle\chi_v^2\rangle$ (solid line), versus r/l for $D\ll 1$ (a), $D=4.5$ (b), and $D\gg 1$ (c).

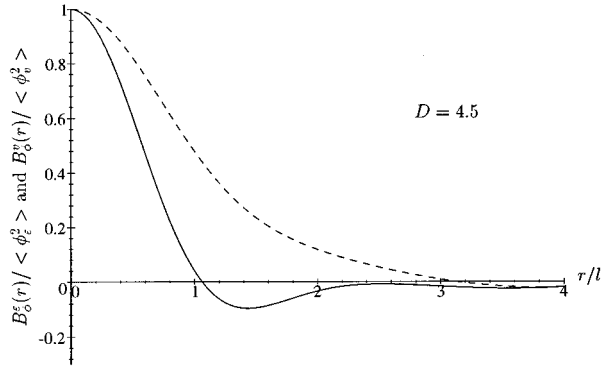


FIG. 2. The normalized temperature and medium velocity contributions to the correlation function for phase fluctuations, $B_\phi^\epsilon(r)/\langle\phi_\epsilon^2\rangle$ (dashed line) and $B_\phi^v(r)/\langle\phi_v^2\rangle$ (solid line), versus r/l for $D=4.5$.

$$\langle\chi^2\rangle=\langle\chi_\epsilon^2\rangle+\langle\chi_v^2\rangle=\frac{\sqrt{\pi}k^2lx}{8}\left[\left(1-\frac{\arctan D}{D}\right)\sigma_\epsilon^2+\left(1-\frac{1}{1+D^2}\right)\frac{4\sigma_v^2}{c_0^2}\right], \tag{18}$$

$$\langle\phi^2\rangle=\langle\phi_\epsilon^2\rangle+\langle\phi_v^2\rangle=\frac{\sqrt{\pi}k^2lx}{8}\left[\left(1+\frac{\arctan D}{D}\right)\sigma_\epsilon^2+\left(1+\frac{1}{1+D^2}\right)\frac{4\sigma_v^2}{c_0^2}\right], \tag{19}$$

where $\langle\chi_\epsilon^2\rangle$ and $\langle\chi_v^2\rangle$ are the temperature and medium velocity contributions to $\langle\chi^2\rangle$, and $\langle\phi_\epsilon^2\rangle$ and $\langle\phi_v^2\rangle$ are the temperature and medium velocity contributions to $\langle\phi^2\rangle$.

Using Eqs. (9), (10), (14), (15), and (17)–(19), the normalized temperature and medium velocity contributions to the correlation function for log-amplitude fluctuations, $B_\chi^\epsilon(r)/\langle\chi_\epsilon^2\rangle$ and $B_\chi^v(r)/\langle\chi_v^2\rangle$, are plotted in Fig. 1 for $D\ll1$, $D=4.5$, and $D\gg1$. These normalized contributions are only functions of r/l and D. The difference in B_χ^ϵ and B_χ^v and, hence, the difference in the temperature and medium velocity contributions to B_χ are clearly seen in Fig. 1 for various values of r/l and D. The normalized temperature and medium velocity contributions to the correlation function for phase fluctuations, $B_\phi^\epsilon(r)/\langle\phi_\epsilon^2\rangle$ and $B_\phi^v(r)/\langle\phi_v^2\rangle$, are represented in Fig. 2 for $D=4.5$. The difference between B_ϕ^ϵ and B_ϕ^v is also clearly seen from the figure. Note that it follows from Eqs. (15) and (17) that if $D\ll1$ or $D\gg1$, the functions $B_\phi^\epsilon(r)/\langle\phi_\epsilon^2\rangle$ and $B_\phi^v(r)/\langle\phi_v^2\rangle$ become the functions $B_\chi^\epsilon(r)/\langle\chi_\epsilon^2\rangle$ and $B_\chi^v(r)/\langle\chi_v^2\rangle$, respectively, calculated for the case $D\gg1$. The latter two functions are plotted in Fig. 1(c).

The correlation radii r_χ^ϵ, r_χ^v, and r_ϕ^ϵ, and r_ϕ^v of the correlation functions $B_\chi^\epsilon(r)$, $B_\chi^v(r)$, $B_\phi^\epsilon(r)$, and $B_\phi^v(r)$, respectively, are determined by

$$B_\chi^\epsilon(r_\chi^\epsilon)=e^{-1}B_\chi^\epsilon(0),\quad B_\chi^v(r_\chi^v)=e^{-1}B_\chi^v(0),\qquad B_\phi^\epsilon(r_\phi^\epsilon)=e^{-1}B_\phi^\epsilon(0),\quad B_\phi^v(r_\phi^v)=e^{-1}B_\phi^v(0). \tag{20}$$

The correlation radii are important parameters characterizing the distance in the plane perpendicular to the direction of sound propagation, at which the correlation of log-amplitude or phase fluctuations is reduced by a factor $1/e$. The correlation radii are calculated numerically on the basis of Eqs. (9), (10), and (20). The normalized correlation radii r_χ^ϵ/l, r_χ^v/l, r_ϕ^ϵ/l, and r_ϕ^v/l are plotted in Fig. 3. All of the normalized correlation radii are only functions of D and are of the order of 1. It follows from the figure that r_χ^ϵ and r_ϕ^ϵ are substantially greater than r_χ^v and r_ϕ^v, respectively. The maximum of the ratio r_χ^ϵ/r_χ^v is 1.52 for $D\to\infty$. The maximum of the ratio r_ϕ^ϵ/r_ϕ^v is 1.75 for $D=4.53$, and, in accordance with Eq. (17), the ratio r_ϕ^ϵ/r_ϕ^v is also 1.52 for $D\to\infty$. Moreover, r_ϕ^ϵ and r_ϕ^v have maxima at $D=3.34$ and $D=1.38$, respectively, while r_χ^ϵ and r_χ^v are monotonically increasing functions of D.

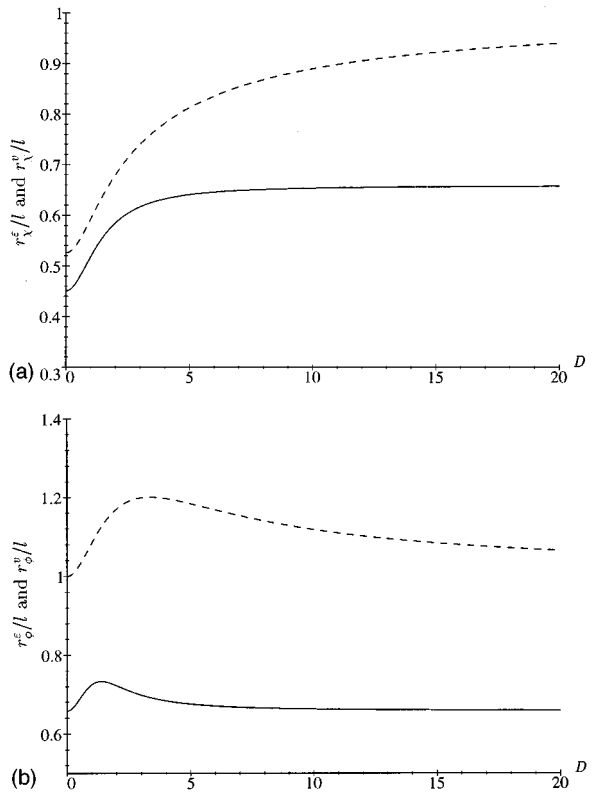


FIG. 3. The normalized correlation radii versus D: r_χ^ϵ/l (dashed line) and r_χ^v/l (solid line) (a), r_ϕ^ϵ/l (dashed line) and r_ϕ^v/l (solid line) (b).

It follows from the theories of waves in random media[6,15] that the correlation radii of $B_\chi(r)$ and $B_\phi(r)$ are of the order of the correlation radius of medium inhomogeneities. For the Gaussian correlation functions of temperature and medium velocity fluctuations, the correlation function $B_{\rm eff}(x,\mathbf{r})$ of $\epsilon_{\rm mov}$ is given by[2]

$$B_{\rm eff}(x,\mathbf{r})=\left[\sigma_\epsilon^2+\left(1-\frac{r^2}{l^2}\right)\frac{4\sigma_v^2}{c_0^2}\right]e^{-R^2/l^2}. \tag{21}$$

The functions $B_{\rm eff}(x,\mathbf{r})$ and $\Phi_{\rm eff}(K_x,\mathbf{K}_r)$ are related by a Fourier transformation similar to Eq. (2). It can be shown from Eq. (21) that the correlation radius l_ϵ of temperature fluctuations is equal to l, while the correlation radius l_v of medium velocity fluctuations is given by $0.658l$. Thus $l_\epsilon>l_v$. Namely due to this inequality the dashed lines in Fig. 3(a) and (b), corresponding to the case of sound propagation in a

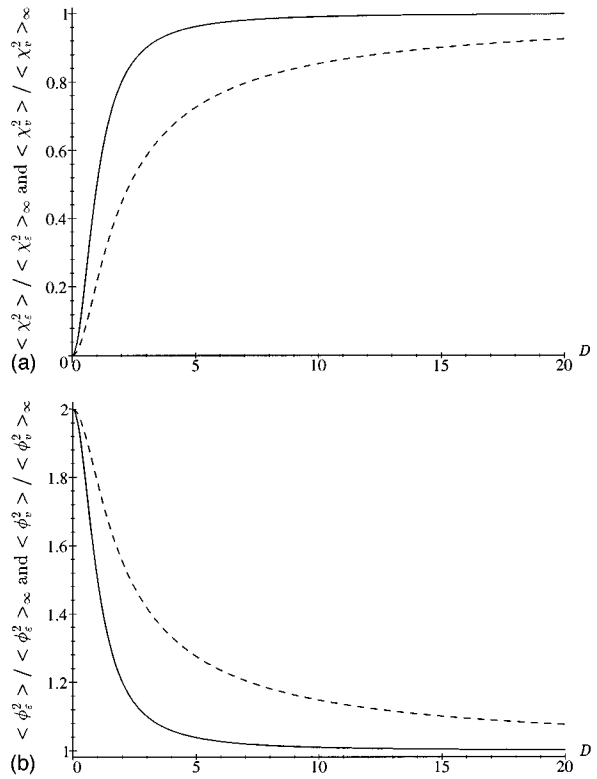


FIG. 4. The normalized temperature and medium velocity contributions to the variances of log-amplitude and phase fluctuations versus D: $\langle\chi_\epsilon^2\rangle/\langle\chi_\epsilon^2\rangle_\infty$ (dashed line) and $\langle\chi_v^2\rangle/\langle\chi_v^2\rangle_\infty$ (solid line) (a); $\langle\phi_\epsilon^2\rangle/\langle\phi_\epsilon^2\rangle_\infty$ (dashed line) and $\langle\phi_v^2\rangle/\langle\phi_v^2\rangle_\infty$ (solid line) (b).

medium with temperature fluctuations, are located above the solid ones corresponding to sound propagation in a medium with medium velocity fluctuations.

E. Variances of log-amplitude and phase fluctuations

The variances of log-amplitude and phase fluctuations are given by Eqs. (18) and (19). Let us consider the normalized temperature and medium velocity contributions to the variances: $\langle\chi_\epsilon^2\rangle/\langle\chi_\epsilon^2\rangle_\infty$, $\langle\chi_v^2\rangle/\langle\chi_v^2\rangle_\infty$, $\langle\phi_\epsilon^2\rangle/\langle\phi_\epsilon^2\rangle_\infty$, and $\langle\phi_v^2\rangle/\langle\phi_v^2\rangle_\infty$. Here, $\langle\chi_\epsilon^2\rangle_\infty=\langle\phi_\epsilon^2\rangle_\infty=\sqrt{\pi}k^2lx\sigma_\epsilon^2/8$ and $\langle\chi_v^2\rangle_\infty=\langle\phi_v^2\rangle_\infty=\sqrt{\pi}k^2lx\sigma_v^2/2c_0^2$ are the variances for $D\to\infty$. Note that $\langle\chi_\epsilon^2\rangle_\infty=\langle\phi_\epsilon^2\rangle_\infty=\langle\chi_v^2\rangle_\infty=\langle\phi_v^2\rangle_\infty$ if Eq. (12) is valid. The normalized temperature and medium velocity contributions to $\langle\chi^2\rangle$ and $\langle\phi^2\rangle$ only depend on D and are given by the corresponding expressions in the parentheses of Eqs. (18) and (19). They are plotted in Fig. 4. It follows from the figure that $\langle\chi_v^2\rangle$ is greater than $\langle\chi_\epsilon^2\rangle$ for all values of D. On the other hand, $\langle\phi_v^2\rangle$ is less than $\langle\phi_\epsilon^2\rangle$ for all values of D in this case. It can be shown from Eqs. (18) and (19) that the maximum of the ratio $(\langle\chi_v^2\rangle/\langle\chi_v^2\rangle_\infty)/(\langle\chi_\epsilon^2\rangle/\langle\chi_\epsilon^2\rangle_\infty)$ is 3 for $D=0$, while the ratio $(\langle\phi_v^2\rangle/\langle\phi_v^2\rangle_\infty)/(\langle\phi_\epsilon^2\rangle/\langle\phi_\epsilon^2\rangle_\infty)$ has a minimum equal to 0.77 at $D=2.31$. If $D\to\infty$, both ratios tend to 1. In other words, the temperature and medium velocity contributions to $\langle\chi^2\rangle$ and $\langle\phi^2\rangle$ are the same for large enough values of D. On the other hand, these contributions are different in the case of the correlation functions $B_\chi(r)$ and $B_\phi(r)$ even for $D\gg1$; see Fig. 1.

F. Attenuation coefficient

Substituting $\Phi_{\text{eff}}(0,K)$ from Eq. (8) into Eq. (6) and calculating the integral over K, we obtain the attenuation coefficient γ of the mean sound field

$$\gamma=\gamma_\epsilon+\gamma_v=\frac{\sqrt{\pi}k^2l}{8}\left(\sigma_\epsilon^2+\frac{4\sigma_v^2}{c_0^2}\right). \tag{22}$$

Here, γ_ϵ and γ_v are the temperature and medium velocity contributions to γ. The formula for γ_ϵ is well-known in the literature,[15] while that for γ_v is a new result. It follows from Eq. (22) that $\gamma_\epsilon=\gamma_v$ if Eq. (12) is valid. γ is a unique statistical characteristic of a sound field, calculated in this paper, which has the same contribution from temperature and medium velocity fluctuations.

G. Coherence function

We substitute $\Phi_{\text{eff}}(0,K)$ from Eq. (8) into Eq. (7), calculate the integral over K, and express the coherence function in the form $\Gamma(r)=I_0\Gamma_\epsilon(r)\Gamma_v(r)$. Here, Γ_ϵ and Γ_v are the factors of Γ due to sound scattering by temperature and medium velocity fluctuations, given by

$$\Gamma_\epsilon(r)=\exp\{-2\gamma_\epsilon x(1-e^{-r^2/l^2})\}, \tag{23}$$

$$\Gamma_v(r)=\exp\left\{-2\gamma_v x\left[1-\left(1-\frac{r^2}{l^2}\right)e^{-r^2/l^2}\right]\right\}. \tag{24}$$

The formula for Γ_ϵ is known in the literature,[15] while that for Γ_v is a new result. The difference in Γ_ϵ and Γ_v is evident from these formulas.

The coherence radius r_Γ^ϵ of the sound field in a motionless random medium is an important statistical characteristic, determined by $\Gamma_\epsilon(r_\Gamma^\epsilon)=e^{-1}$.[6] This definition is meaningful if $\gamma_\epsilon x>1/2$. Indeed, $\Gamma_\epsilon(r)>e^{-1}$ if $\gamma_\epsilon x<1/2$. It can be shown from Eq. (23) that

$$r_\Gamma^\epsilon=l\ \ln^{1/2}\left(\frac{1}{1-1/(2\gamma_\epsilon x)}\right). \tag{25}$$

For $\gamma_\epsilon x\gg1$, this formula becomes $r_\Gamma^\epsilon/l\simeq(2\gamma_\epsilon x)^{-1/2}$. The coherence radius r_Γ^v of the sound field in a moving random medium is determined by analogous equation $\Gamma_v(r_\Gamma^v)=e^{-1}$. From this definition and Eq. (24) it follows that r_Γ^v is a solution of the equation

$$2\gamma_v x\left[1-\left(1-\frac{r^2}{l^2}\right)e^{-r^2/l^2}\right]=1. \tag{26}$$

For $\gamma_v x\gg1$, Eq. (26) has the following approximate solution: $r_\Gamma^v/l\simeq(4\gamma_v x)^{-1/2}$.

The coherence radii r_Γ^ϵ and r_Γ^v, normalized by l, are plotted in Fig. 5 as functions of $\gamma_\epsilon x=\gamma_v x$. The coherence radius r_Γ^ϵ is greater than r_Γ^v for all values of the argument. Moreover, r_Γ^ϵ tends to infinity if $\gamma_\epsilon x=\gamma_v x\to1/2$, while the limit of r_Γ^v/l is equal to 1 for this case.

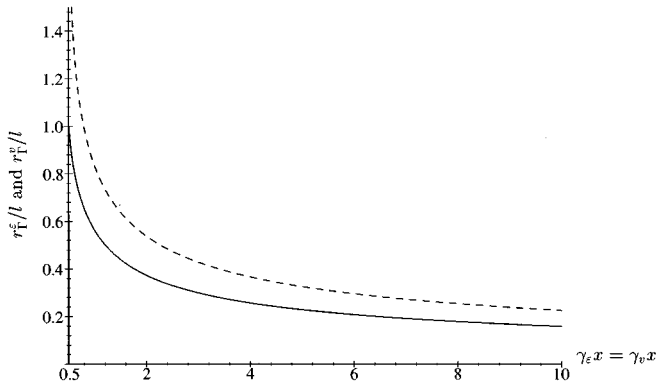


FIG. 5. The normalized coherence radii r_Γ^ϵ/l (dashed line) and r_Γ^v/l (solid line) versus $\gamma_\epsilon x = \gamma_v x$.

H. Low-frequency sound propagation

So far in this paper we have considered the case where the scale L_0 of largest inhomogeneities in a medium, which affect sound propagation, is greater than the wavelength λ. This inequality may be violated for low enough frequencies. In this subsection, we shall consider the case when the ratio between L_0 and λ can be arbitrary. For this case, the mean sound field $\langle p \rangle$, calculated by using the Feynman diagram technique and Bourret approximation,[6] is still given by Eq. (5) if the attenuation coefficient γ is given by the formula[2]

$$\gamma = \frac{\pi^2 k^2}{2} \int_0^{2k} K \left[\left(\beta_c + \frac{K^2 \beta_\varrho}{2k^2} \right)^2 \frac{\Phi_T(K)}{T_0^2} + 4 \left(1 - \frac{K^2}{2k^2} \right)^2 \left(1 - \frac{K^2}{4k^2} \right) \frac{F(K)}{c_0^2} \right] dK, \tag{27}$$

which differs from Eq. (6). [As it should be, Eq. (27) coincides with Eq. (6) if $L_0 \gg \lambda$.] In Eq. (27), $\Phi_T(K)$ and $F(K)$ are the three-dimensional spectral densities of the random fields $\tilde{T}$ and $\mathbf{v}=(v_x, v_y, v_z)$, respectively, the coefficient $\beta_\varrho = (T_0/\varrho_0)(\partial \varrho_0 / \partial T_0)$, where ϱ_0 is the mean value of the medium density ϱ. Note that if $L_0 < \lambda$, the sound field is affected not only by the random fields ϵ and $\mathbf{v}$, but also by fluctuations in the density $\tilde{\varrho} = \varrho - \varrho_0$. Therefore, when deriving Eq. (27), it is worthwhile expressing the random fields ϵ and $\tilde{\varrho}$ through the temperature fluctuations by the formulas $\epsilon = -\beta_c \tilde{T}/T_0$ and $\tilde{\varrho} = \varrho_0 \beta_\varrho \tilde{T}/T_0$.

For the assumed Gaussian spectra of temperature and medium velocity fluctuations, Φ_T and F are given by[2]

$$\Phi_T(K) = (\sigma_T^2 l^3 / 8\pi^{3/2}) \exp(-K^2 l^2/4),$$
$$F(K) = (\sigma_v^2 K^2 l^5 / 32\pi^{3/2}) \exp(-K^2 l^2/4). \tag{28}$$

The main difference between Φ_T and F is the factor K^2 before the exponential term in Eq. (28), similar to the factor $K^2 l^2/4$ in Eq. (8). It is well-known in the theory of turbulence[3] that this factor must appear in a three-dimensional spectral density of the solenoidal vector random field $\mathbf{v}$. Substituting Eq. (28) into Eq. (27), integrating over K, and assuming that $\beta_c = -\beta_\varrho = 1$ (these equalities are valid for sound propagation in the atmosphere), yields the desired formula for the attenuation coefficient

$$\gamma = \gamma_T + \gamma_v = \frac{\sqrt{\pi} k^2 l}{8} \left\{ \left[1 - \frac{4}{k^2 l^2} + \frac{8}{k^4 l^4} - \left(1 + \frac{4}{k^2 l^2} + \frac{8}{k^4 l^4} \right) e^{-k^2 l^2} \right] \frac{\sigma_T^2}{T_0^2} + \left[1 - \frac{10}{k^2 l^2} + \frac{48}{k^4 l^4} - \frac{96}{k^6 l^6} + \left(1 + \frac{10}{k^2 l^2} + \frac{48}{k^4 l^4} + \frac{96}{k^6 l^6} \right) e^{-k^2 l^2} \right] \frac{4\sigma_v^2}{c_0^2} \right\}. \tag{29}$$

Here, γ_T and γ_v are the temperature and medium velocity contributions to γ. Equation (29) is a new result obtained in this paper. The difference in γ_T and γ_v, i.e., in the temperature and medium velocity contributions to γ, is seen clearly from this equation.

For $kl \ll 1$, Eq. (29) becomes

$$\gamma = \gamma_T + \gamma_v = \frac{\sqrt{\pi} k^4 l^3}{24} \left(\frac{\sigma_T^2}{T_0^2} + \frac{k^2 l^2}{10} \frac{4\sigma_v^2}{c_0^2} \right). \tag{30}$$

It follows from this formula that $\gamma_T \gg \gamma_v$ if Eq. (12) is valid. Here, we take into account that $\sigma_\epsilon = \sigma_T^2/T_0^2$ for the considered case $\beta_c = -1$. On the other hand, for $kl \gg 1$, Eq. (29) becomes

$$\gamma = \gamma_T + \gamma_v = \frac{\sqrt{\pi} k^2 l}{8} \left(\frac{\sigma_T^2}{T_0^2} + \frac{4\sigma_v^2}{c_0^2} \right). \tag{31}$$

It follows from this formula that $\gamma_T = \gamma_v$ if Eq. (12) is valid. Equation (31) coincides with Eq. (22) since $\sigma_\epsilon^2 = \sigma_T^2/T_0^2$.

II. SOUND SCATTERING CROSS SECTION

The sound scattering cross section per unit volume, σ, is one of the most important statistical characteristics of a sound field in a random medium. The formula for σ in a statistically isotropic moving random medium is given by[2]

$$\sigma(\theta) = \frac{\pi k^4}{2} [\beta^2(\theta) \Phi_T(2k \sin(\theta/2))/T_0^2 + 4 \cos^2 \theta \cos^2(\theta/2) F(2k \sin(\theta/2))/c_0^2]. \tag{32}$$

Here, θ is the scattering angle, and the coefficient $\beta(\theta) = \beta_c + 2\beta_\varrho \sin^2(\theta/2)$ is a linear combination of the coefficients β_c and β_ϱ, introduced above. Equation (32) for σ does not depend on the form of a sound wave incident on a scattering volume and is, therefore, valid both for plane and spherical waves. For the Gaussian spectra of temperature and medium velocity fluctuations, Φ_T and F are given by Eq. (28). Substituting these equations into Eq. (32) yields the desired formula for σ:

$$\sigma(\theta)=\sigma^T(\theta)+\sigma^v(\theta)$$

$$=\frac{k^4l^3}{16\sqrt{\pi}}\,e^{-k^2l^2\sin^2(\theta/2)}$$

$$\times\left[\beta^2(\theta)\,\frac{\sigma_T^2}{T_0^2}+k^2l^2\,\frac{\sigma_v^2}{c_0^2}\cos^2\theta\,\sin^2\theta\right]. \tag{33}$$

Here, σ^T and σ^v are the temperature and medium velocity contributions to σ. We shall compare these contributions for the case of sound scattering in the atmosphere, where $\beta^2(\theta)=\cos^2\theta$. It follows from Eq. (33) that σ^T and σ^v differ not only quantitatively but also qualitatively. First, σ^T is proportional to k^4, while $\sigma^v\sim k^6$. Second, σ^T and σ^v have different dependences on the scattering angle θ: σ^T becomes zero only at $\theta=\pi$, while σ_v becomes zero at $\theta=0$, π, and 2π.

If $kl\gg1$, the scattering is only significant within a cone with the axis in the direction of the incident wave and with the angle $\theta\sim1/kl$. In this case, Eq. (33) becomes

$$\sigma(\theta)=\frac{k^4l^3}{16\sqrt{\pi}}\,e^{-k^2l^2\theta^2/4}\left[\frac{\sigma_T^2}{T_0^2}+k^2l^2\theta^2\,\frac{\sigma_v^2}{c_0^2}\right]. \tag{34}$$

In the other limiting case $kl\ll1$, σ^v may be ignored in comparison with σ^T in Eq. (33).

III. CONCLUSIONS

The formulas for the most often used statistical moments of a plane wave propagating in a moving random medium with arbitrary (anisotropic) spectra of temperature and medium velocity fluctuations have been obtained. Then, these statistical moments have been calculated for the Gaussian correlation functions of temperature and medium velocity fluctuations. The temperature and medium velocity contributions to these statistical moments have been compared in detail. It has been shown that these contributions may differ not only quantitatively but also qualitatively. This is explained by the fact that temperature fluctuations are a scalar random field while medium velocity fluctuations are a vector random field, and these random fields affect sound propagation in a different manner. The obtained results may be used in atmospheric and oceanic acoustics and aeroacoustics for investigating sound propagation through a moving random medium. At present, we do not have experimental data on plane wave propagation in a medium with the Gaussian spectrum of medium velocity fluctuations, to which the obtained theoretical results can be compared. Therefore, it is worthwhile noting that the general theory described in the Introduction and Sec. I A has also been used in Ref. 20 for analytical calculations of the probability density of distance to the first caustic for plane wave propagation in a moving random medium with the Gaussian spectrum. The analytical calculations of the probability density of distance are shown to be in a good agreement with numerical simulations.[20] That supports the validity of the general theory, which is used for deriving the new results obtained in the present paper.

ACKNOWLEDGMENT

This material is based on work supported in part by the U.S. Army Research Office under Contract No. DAAH04-95-1-0593.

APPENDIX: CALCULATIONS OF THE CORRELATION FUNCTIONS OF LOG-AMPLITUDE AND PHASE FLUCTUATIONS

In this Appendix we present the derivation of Eqs. (9) and (10) for the correlation functions $B_\chi(r)$ and $B_\phi(r)$. In Eq. (4), we replace $\Phi_{\rm eff}(0,K)$ by its value given by Eq. (8). This yields the formulas for B_χ and B_ϕ in a moving random medium

$$B_{\chi,\phi}(r)=\frac{\sqrt{\pi}k^2l^3x}{16}\int_0^\infty\left\{\left[K\mp\frac{K_F^2}{K}\sin\frac{K^2}{K_F^2}\right]\sigma_\epsilon^2+\left[\frac{K^3l^2}{c_0^2}\right.\right.$$

$$\left.\left.\mp l^2K_F^2K\sin\frac{K^2}{K_F^2}\right]\frac{\sigma_v^2}{c_0^2}\right\}e^{-k^2l^2/4}J_0(Kr)dK. \tag{A1}$$

The upper and lower signs on the right-hand side of Eq. (A1) correspond to B_χ and B_ϕ, respectively. Equation (A1) can be expressed in the form

$$B_{\chi,\phi}(r)=\frac{\sqrt{\pi}k^2l^3x}{16}\left[(I_1\mp I_2)\sigma_\epsilon^2+(I_3\mp I_4)\,\frac{\sigma_v^2}{c_0^2}\right], \tag{A2}$$

where the integrals I_{1-4} are given by

$$I_1=\int_0^\infty K\,\exp\left(-\frac{K^2l^2}{4}\right)J_0(Kr)dK, \tag{A3}$$

$$I_2=K_F^2\Im\int_0^\infty K^{-1}\exp\left[-K^2\left(\frac{l^2}{4}-\frac{i}{K_F^2}\right)\right]J_0(Kr)dK, \tag{A4}$$

$$I_3=l^2\int_0^\infty K^3\,\exp\left(-\frac{K^2l^2}{4}\right)J_0(Kr)dK, \tag{A5}$$

$$I_4=K_F^2l^2\Im\int_0^\infty K\,\exp\left[-K^2\left(\frac{l^2}{4}-\frac{i}{K_F^2}\right)\right]J_0(Kr)dK. \tag{A6}$$

In Eqs. (A3)–(A6), the identity $\Im e^{iy}=\sin y$ is used. The integrals I_{1-4} can be calculated by making use of the reference integral[21]

$$I=\int_0^\infty K^{\alpha-1}e^{-qK^2}J_0(Kr)dK$$

$$=q^{-\alpha/2}2^{-1}\Gamma\left(\frac{\alpha}{2}\right)\,{}_1F_1\left(\frac{\alpha}{2};1;\frac{-r^2}{4q}\right), \tag{A7}$$

where Γ is the Gamma function, ${}_1F_1$ is the confluent hypergeometric function, α and q are parameters, and $\Re\alpha>0$. When calculating I_1, we use Eq. (A7) with $\alpha=2$ and take into account that

$${}_1F_1(1;1;y)=\exp(y). \tag{A8}$$

As a result, we obtain

$$I_1=(2/l^2)\exp(-r^2/l^2). \tag{A9}$$

To calculate I_3, we set $\alpha=4$ in Eq. (A7) and take into account that ${}_1F_1(2;1;y)=(1+y)\exp(y)$. This yields

$$I_3=(8/l^2)(1-r^2/l^2)\exp(-r^2/l^2). \tag{A10}$$

Using Eq. (A7) with $\alpha=2$ yields the formula for I_4

$$I_4=2K_F^2 l^2 \Im\left[\left(l^2-\frac{4i}{K_F^2}\right)^{-1} {}_1F_1\left(1;1;-\frac{r^2}{l^2-4i/K_F^2}\right)\right]. \tag{A11}$$

In this formula, ${}_1F_1$ can be replaced by an exponent by using Eq. (A8). Then, taking the imaginary part of the square brackets in the resulting formula yields

$$I_4=\frac{8\exp(-r^2/(l^2(1+D^2)))}{l^2(1+D^2)}\left[\cos\left(\frac{r^2D}{l^2(1+D^2)}\right)-\frac{1}{D}\sin\left(\frac{r^2D}{l^2(1+D^2)}\right)\right]. \tag{A12}$$

The calculation of I_2 is more complicated. Equation (A7) cannot be used directly to calculate I_2 since $\alpha=0$ in Eq. (A4) for I_2. Therefore, we first represent this equation in the form

$$I_2=K_F^2 \lim_{\nu\to+0} \Im\int_0^\infty K^{2\nu-1} \times\exp\left[-K^2\left(\frac{l^2}{4}-\frac{i}{K_F^2}\right)\right]J_0(Kr)\,dK. \tag{A13}$$

Now Eq. (A7) can be used to calculate the integral in Eq. (A13). The result is

$$I_2=\frac{K_F^2}{2}\lim_{\nu\to+0}\Im\left[\left(\frac{l}{2}\right)^{-2\nu}(1-iD)^{-\nu}\Gamma(\nu) \times{}_1F_1\left(\nu;1;-\frac{r^2}{l^2(1-iD)}\right)\right]. \tag{A14}$$

Using the identities $\Gamma(1+\nu)=\nu\Gamma(\nu)$, and $(1-iD)^\nu=(1+D^2)^{\nu/2}e^{-i\varphi\nu}$, where $\varphi=\arctan D$, Eq. (A14) can be represented in the form

$$I_2=(K_F^2/2)\lim_{\nu\to+0}\Im[\Omega(\nu)\nu^{-1}e^{i\varphi\nu}{}_1F_1(\nu;1;y)]. \tag{A15}$$

Here, $y=-r^2\exp(i\varphi)/(l^2\sqrt{1+D^2})$, and $\Omega(\nu)=(l/2)^{-2\nu}(1+D^2)^{-\nu/2}\Gamma(1+\nu)$. Since $\Omega(\nu)$ is a real function and has a limit equal to 1 as $\nu\to+0$, it can be replaced by the value of this limit in Eq. (A15). Using the formula $\exp(i\varphi\nu)=\cos(\varphi\nu)+i\sin(\varphi\nu)$ and representing ${}_1F_1(\nu;1;y)$ in the form of a series in y yields

$$I_2=\frac{K_F^2}{2}\lim_{\nu\to+0}\Im\left[\frac{\cos(\varphi\nu)}{\nu}+\frac{i\sin(\varphi\nu)}{\nu} +(\cos(\varphi\nu)+i\sin(\varphi\nu))\left(y+\frac{(\nu+1)y^2}{2!2!}+\frac{(\nu+1)(\nu+2)y^3}{3!3!}+\cdots\right)\right]. \tag{A16}$$

The imaginary part of the first term in the square brackets in Eq. (A16) is zero and, therefore, it can be omitted. After omitting this term, the order of the symbols lim and $\Im$ can be changed, and the limit can be calculated

$$I_2=\frac{K_F^2}{2}\Im\lim_{\nu\to+0}\left[\frac{i\sin(\varphi\nu)}{\nu}+(\cos(\varphi\nu)+i\sin(\varphi\nu)) \times\left(y+\frac{(\nu+1)y^2}{2!2!}+\frac{(\nu+1)(\nu+2)y^3}{3!3!}+\cdots\right)\right] =\frac{K_F^2}{2}\left[\varphi+\Im\left(y+\frac{y^2}{2!2}+\frac{y^3}{3!3}+\cdots\right)\right]. \tag{A17}$$

The series in parentheses on the right-hand side of Eq. (A17) is equal to $-\gamma_E-\ln(-y)-E_1(-y)$, where $\gamma_E=0.577$ is the Euler constant.[22] Taking into account that $\Im\ln(-y)=\varphi$, one obtains the final formula for I_2:

$$I_2=-\frac{K_F^2}{2}\Im E_1\left(\frac{r^2\exp(i\varphi)}{l^2\sqrt{1+D^2}}\right). \tag{A18}$$

Substituting the obtained formulas for I_{1-4} into Eq. (A2) yields Eqs. (9) and (10) for B_χ and B_ϕ.

[1] V. E. Ostashev, *Sound Propagation in Moving Media* (Nauka, Moscow, 1992) (in Russian).
[2] V. E. Ostashev, ''Sound propagation and scattering in media with random inhomogeneities of sound speed, density and medium velocity,'' Waves Random Media **4**, 403–428 (1994).
[3] J. Hinze, *Turbulence* (McGraw-Hill, New York, 1975).
[4] V. I. Tatarskii, *The Effects of the Turbulent Atmosphere on Wave Propagation* (Israel Program for Scientific Translation, Jerusalem, 1971).
[5] E. H. Brown and F. F. Hall, ''Advances in atmospheric acoustics,'' Rev. Geophys. Space Phys. **16**, 47–110 (1978).
[6] S. M. Rytov, Yu. A. Kravtzov, and V. I. Tatarskii, *Principles of Statistical Radio Physics. Part 4. Wave Propagation through Random Media* (Springer-Verlag, Berlin, 1989).
[7] G. A. Daigle, J. E. Piercy, and T. F. W. Embleton, ''Effects of atmospheric turbulence on the interference of sound waves near a hard boundary,'' J. Acoust. Soc. Am. **64**, 622–630 (1978).
[8] G. A. Daigle, J. E. Piercy, and T. F. W. Embleton, ''Line-of-sight propagation through atmospheric turbulence near the ground,'' J. Acoust. Soc. Am. **74**, 1505–1513 (1983).
[9] G. A. Daigle, T. F. W. Embleton, and J. E. Piercy, ''Propagation of sound in the presence of gradients and turbulence near the ground,'' J. Acoust. Soc. Am. **79**, 613–627 (1986).
[10] M. A. Johnson, R. Raspet, and M. T. Bobak, ''A turbulence model for sound propagation from an elevated source above level ground,'' J. Acoust. Soc. Am. **81**, 638–646 (1987).
[11] W. E. McBride, H. E. Bass, R. Raspet, and K. E. Gilbert, ''Scattering of sound by atmospheric turbulence: A numerical simulation above a complex impedance boundary,'' J. Acoust. Soc. Am. **90**, 3314–3325 (1991).
[12] W. E. McBride, H. E. Bass, R. Raspet, and K. E. Gilbert, ''Scattering of sound by atmospheric turbulence: Predictions in the refractive shadow zone,'' J. Acoust. Soc. Am. **91**, 1336–1340 (1992).
[13] E. M. Salomons, ''Downwind propagation of sound in an atmosphere with a realistic sound-speed profile: A semianalytical ray model,'' J. Acoust. Soc. Am. **95**, 2425–2436 (1994).
[14] L. A. Chernov, *Wave Propagation in a Random Medium* (Dover, New York, 1960).
[15] L. A. Chernov, *Waves in Randomly Inhomogeneous Media* (Nauka, Moscow, 1975) (in Russian).
[16] D. K. Wilson and D. W. Thomson, ''Acoustic propagation through anisotropic, surface-layer turbulence,'' J. Acoust. Soc. Am. **96**, 1080–1095 (1994).
[17] V. E. Ostashev, ''Propagation and scattering of sound waves in turbulent media (the atmosphere and ocean),'' Atmos. Opt. **4**, 931–937 (1991).
[18] D. Di Iorio and D. M. Farmer, ''Two-dimensional angle of arrival fluctuations,'' J. Acoust. Soc. Am. **100**, 814–824 (1996).

[19] V. E. Ostashev, F. Gerdes, V. Mellert, and R. Wandelt, ''Propagation of sound in a turbulent medium. II. Spherical waves,'' J. Acoust. Soc. Am. **102**, 2571–2578 (1997).
[20] Ph. Blanc-Benon, D. Juve, V. E. Ostashev, and R. Wandelt, ''On the appearance of caustics for plane sound-wave propagation in moving random media,'' Waves Random Media **5**, 183–199 (1995).
[21] A. P. Prudnikov, Y. A. Brychkov, and O. I. Marichev, *Integrals and Series (Elementary Functions)* (Moscow, Nauka, 1981) (in Russian).
[22] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (U.S. Government Printing Office, Washington, DC, 1964).

Propagation of sound in a turbulent medium. II. Spherical waves

V. E. Ostashev[a)]
Department of Physics, New Mexico State University, Las Cruces, New Mexico 88003-8001

Frank Gerdes, Volker Mellert, and Ralf Wandelt
Arbeitsgruppe Akustik, Fachbereich Physik, Universität Oldenburg, Oldenburg D-26111, Germany

(Received 28 March 1996; revised 9 June 1997; accepted for publication 7 July 1997)

Formulas for the statistical moments of a spherical sound wave propagating in a medium with arbitrary (anisotropic) spectra of temperature and medium velocity fluctuations are obtained. These statistical moments are: the variances of log-amplitude and phase fluctuations, the correlation functions of log-amplitude and phase fluctuations, the mean sound field, and the coherence function of the sound field. Then, the statistical moments of a spherical sound wave are calculated analytically and numerically for Gaussian spectra of temperature and medium velocity fluctuations. It is shown that the temperature and medium velocity contributions to these statistical moments can differ not only quantitatively but also qualitatively. The useful relationships between the statistical moments of spherical and plane waves propagating in a moving random medium are derived. Some of the theoretical results obtained are compared with experimental data from the literature on sound propagation through the turbulent atmosphere. © *1997 Acoustical Society of America.* [S0001-4966(97)04810-8]

PACS numbers: 43.28.Fp, 43.20.Bi, 43.20.Fn [LCS]

INTRODUCTION

The propagation of a plane acoustical wave in a moving random medium with temperature and medium velocity fluctuations has been considered in Ref. 1, which is referred hereinafter as Paper I. Formulas for the variances of log-amplitude and phase fluctuations of a sound field, $\langle \chi^2 \rangle$ and $\langle \phi^2 \rangle$, the correlation functions of log-amplitude and phase fluctuations, B_χ and B_ϕ, the mean sound field $\langle p \rangle$, the coherence function Γ, and the sound scattering cross section per unit volume σ have been obtained in Paper I for arbitrary (anisotropic) spectra of medium inhomogeneities. These statistical moments have also been calculated for Gaussian correlation functions for temperature and medium velocity fluctuations.

The present paper is a logical continuation of Paper I. We shall obtain formulas for the aforementioned statistical moments for the case of spherical wave propagation in a moving random medium with arbitrary (anisotropic) spectra of medium inhomogeneities. Then, these statistical moments will be calculated for Gaussian correlation functions for temperature and medium velocity fluctuations. We shall also compare the temperature and medium velocity contributions to the statistical moments of a sound field.

The paper is organized as follows. The statistical moments of a spherical sound wave propagating in a moving random medium are calculated analytically and numerically in Sec. I. Section II presents a comparison of some of theoretical results obtained with experimental data from the literature. Conclusions are summarized in Sec. III. The derivation of formulas for B_χ and B_ϕ is presented in the Appendix.

[a)]On leave from Institute of Atmospheric Physics, Moscow 109017, Russia.

I. THEORETICAL AND NUMERICAL RESULTS

A. Anisotropic turbulence

The Introduction and Sec. I A of Paper I summarize the theory which allows one to derive equations for the statistical moments of a sound field propagating in a moving random medium with arbitrary spectra of temperature and medium velocity fluctuations. In this subsection, this theory is used to obtain formulas for $\langle \chi^2 \rangle$, $\langle \phi^2 \rangle$, B_χ, B_ϕ, $\langle p \rangle$, and Γ for a spherical sound wave in such a medium.

The formulas for the correlation functions of log-amplitude and phase fluctuations of a spherical wave in a moving random medium, calculated by the Rytov method, are obtained from formulas for analogous correlation functions in a motionless random medium[2] and by making use of Eq. (3) from Paper I:

$$B_{\chi,\phi}(r) = \frac{\pi^2 k^2 x}{2} \int_0^1 dt \int_0^\infty dK\, K \left(1 \mp \cos\left(\frac{K^2 t(1-t)}{K_F^2} \right) \right) \times J_0(Krt)\Phi_{\text{eff}}(0,K). \tag{1}$$

Here, t is an integration variable, and the upper and lower signs correspond to B_χ and B_ϕ, respectively. Other notations, which are used in Eq. (1) and in the present paper, are the same as those in Paper I. In comparison with Eq. (4) from Paper I for the corresponding correlation functions of a plane wave, Eq. (1) contains the integral over t, which appears due to the integration along the path of spherical wave propagation.

The variances of log-amplitude and phase fluctuations, $\langle \chi^2 \rangle = B_\chi(0)$ and $\langle \phi^2 \rangle = B_\phi(0)$, of a spherical wave in a moving random medium are given by the right-hand side of Eq. (1) with $r=0$.

Using Ref. 3, it can be shown that the mean sound field $\langle p\rangle$ of a spherical wave, calculated by the parabolic equation method or by the Feynman diagram technique,[4] is given by

$$\langle p(x,\mathbf{r})\rangle = \exp(-\gamma x)p_0(x,\mathbf{r}), \tag{2}$$

where $p_0=\exp(ikR)/4\pi R$. The form of Eq. (2) coincides with that for $\langle p\rangle$ for the case of plane wave propagation [see Eq. (5) from Paper I]. However, p_0 in these equations are different. In the parabolic equation approximation, the attenuation coefficient γ in Eq. (2) is given by Eqs. (6) and (22) from Paper I for arbitrary and Gaussian spectra of medium inhomogeneities, respectively. For these spectra, γ in Eq. (2), calculated by the Feynman diagram technique and Bourret approximation, is given by Eqs. (27) and (29) from Paper I.

The formula for the coherence function $\Gamma(\mathbf{r}_1,\mathbf{r}_2)=\langle p(x,\mathbf{r}_1)p^*(x,\mathbf{r}_2)\rangle$ of a spherical wave in a moving random medium, derived by the parabolic equation method, is obtained from the corresponding formula in a motionless medium[2] and by using Eq. (3) from Paper I:

$$\Gamma(\mathbf{r},\mathbf{r}_+)=\frac{1}{(4\pi x)^2}\exp\left\{\frac{ik\mathbf{r}\cdot\mathbf{r}_+}{x}-\pi^2k^2x\int_0^1 dt\int_0^\infty dK \times K[1-J_0(Krt)]\Phi_{\text{eff}}(0,K)\right\}. \tag{3}$$

Here, $\mathbf{r}=\mathbf{r}_1-\mathbf{r}_2$ and $\mathbf{r}_+=(\mathbf{r}_1+\mathbf{r}_2)/2$. Using Eq. (1), Eq. (3) can be expressed in the form

$$\Gamma(\mathbf{r},\mathbf{r}_+)=\frac{1}{(4\pi x)^2}\exp\left\{\frac{ik\mathbf{r}\cdot\mathbf{r}_+}{x}+B_\chi(r)-B_\chi(0)+B_\phi(r)-B_\phi(0)\right\}. \tag{4}$$

This formula gives us a useful relationship between Γ, calculated by the parabolic equation method, and B_χ and B_ϕ, calculated by using the Rytov method.

As mentioned in Paper I, the sound scattering cross section per unit volume σ is given by the same formulas for plane and spherical waves.

The formulas presented above for $\langle\chi^2\rangle$, $\langle\phi^2\rangle$, B_χ, B_ϕ, and Γ are valid for arbitrary (anisotropic) spectra of temperature and medium velocity fluctuations. In the rest of this section, these statistical moments will be calculated for Gaussian spectra of temperature and medium velocity fluctuations. In this case, $\Phi_{\text{eff}}(0,K)$ in Eqs. (1) and (3) is given by Eq. (8) from Paper I.

B. Correlation functions for log-amplitude and phase fluctuations

We replace Φ_{eff} in Eq. (1) by its value given by Eq. (8) from Paper I and calculate the integral over K (this calculation is presented in the Appendix). As a result, one obtains the expression for the correlation function of log-amplitude fluctuations as the sum of the temperature and medium velocity contributions

$$B_\chi(r)=B_\chi^\epsilon(r)+B_\chi^v(r), \tag{5}$$

where

$$B_\chi^\epsilon(r)=\frac{\sqrt{\pi}k^2lx\sigma_\epsilon^2}{8}\int_0^1 dt\left\{e^{-\rho^2}-\frac{e^{-\rho^2/d}}{d}\times\left[\cos\left(\frac{\rho^2D\tau}{d}\right)+D\tau\sin\left(\frac{\rho^2D\tau}{d}\right)\right]\right\}, \tag{6}$$

$$B_\chi^v(r)=\frac{\sqrt{\pi}k^2lx\sigma_v^2}{2c_0^2}\int_0^1 dt\left\{(1-\rho^2)e^{-\rho^2}-\frac{e^{-\rho^2/d}}{d^3}\times\left[(1-D^4\tau^4+\rho^2(3d-4))\cos\left(\frac{\rho^2D\tau}{d}\right)+D\tau(2d+\rho^2(d-4))\sin\left(\frac{\rho^2D\tau}{d}\right)\right]\right\}. \tag{7}$$

Similarly, the expression for the correlation function of phase fluctuations is the sum of the temperature and medium velocity contributions

$$B_\phi(r)=B_\phi^\epsilon(r)+B_\phi^v(r), \tag{8}$$

where

$$B_\phi^\epsilon(r)=\frac{\sqrt{\pi}k^2lx\sigma_\epsilon^2}{8}\int_0^1 dt\left\{e^{-\rho^2}+\frac{e^{-\rho^2/d}}{d}\times\left[\cos\left(\frac{\rho^2D\tau}{d}\right)+D\tau\sin\left(\frac{\rho^2D\tau}{d}\right)\right]\right\}, \tag{9}$$

$$B_\phi^v(r)=\frac{\sqrt{\pi}k^2lx\sigma_v^2}{2c_0^2}\int_0^1 dt\left\{(1-\rho^2)e^{-\rho^2}+\frac{e^{-\rho^2/d}}{d^3}\times\left[(1-D^4\tau^4+\rho^2(3d-4))\cos\left(\frac{\rho^2D\tau}{d}\right)+D\tau(2d+\rho^2(d-4))\sin\left(\frac{\rho^2D\tau}{d}\right)\right]\right\}. \tag{10}$$

In Eqs. (6), (7), (9), and (10), $\rho=rt/l$, $\tau=t(1-t)$, and $d=1+D^2\tau^2$ are new variables, which depend on t and are used to simplify the equations. For the limiting cases $D\ll 1$ and $D\gg 1$, the equations for B_χ^ϵ and B_ϕ^ϵ were obtained in Ref. 5, where the case of temperature fluctuations has been considered only. Equations (7) and (10) for B_χ^v and B_ϕ^v are new results obtained in the present paper.

When comparing the temperature and medium velocity contributions to the statistical moments of a sound field, in this paper as in Paper I, we shall assume that Eq. (12) from Paper I is valid. In this case, it follows from Eqs. (6), (7), (9), and (10) that $B_\chi^\epsilon\neq B_\chi^v$ and $B_\phi^\epsilon\neq B_\phi^v$, i.e., the temperature and medium velocity contributions to the correlation functions B_χ and B_ϕ are different.

There are three limiting cases of the derived Eqs. (5)–(10), which are worth considering. The first limiting case is for $r=0$. In this case, $B_\chi(0)=\langle\chi^2\rangle$ and $B_\phi(0)=\langle\phi^2\rangle$. Setting $r=0$ in Eqs. (5)–(10) and calculating the integrals over t yields formulas for the variances of log-amplitude and phase fluctuations

$$\langle\chi^2\rangle=\langle\chi_\epsilon^2\rangle+\langle\chi_v^2\rangle=\frac{\sqrt{\pi}k^2lx}{8}\left\{[1-F_1(D)]\sigma_\epsilon^2+[1-F_2(D)]\,\frac{4\sigma_v^2}{c_0^2}\right\},\tag{11}$$

$$\langle\phi^2\rangle=\langle\phi_\epsilon^2\rangle+\langle\phi_v^2\rangle=\frac{\sqrt{\pi}k^2lx}{8}\left\{[1+F_1(D)]\sigma_\epsilon^2+[1+F_2(D)]\,\frac{4\sigma_v^2}{c_0^2}\right\}.\tag{12}$$

Here, $\langle\chi_\epsilon^2\rangle$ and $\langle\chi_v^2\rangle$ are the temperature and medium velocity contributions to $\langle\chi^2\rangle$, while $\langle\phi_\epsilon^2\rangle$ and $\langle\phi_v^2\rangle$ are those for $\langle\phi^2\rangle$. Moreover, in Eqs. (11) and (12)

$$F_1(D)=\frac{\arctan\sqrt{2/\Omega}+\dfrac{\Omega\Delta}{2}\ln[(1+\Delta\sqrt{2\Omega})/(1-\Delta\sqrt{2\Omega})]}{\Delta^2(\Omega+1)\sqrt{8\Omega}},\tag{13}$$

$$F_2(D)=\frac{\Omega(\Omega+2)}{2(\Omega+1)^2}\left[1+\frac{\sqrt{2\Omega}(\Omega+3)\arctan\sqrt{2/\Omega}}{4(\Omega+1)}+\frac{\Delta\sqrt{2\Omega}(\Omega-1)(\Omega+2)\ln[(1+\Delta\sqrt{2\Omega})/(1-\Delta\sqrt{2\Omega})]}{8(\Omega+1)}\right],\tag{14}$$

where $\Delta=D/4$ and $\Omega=\sqrt{1+\Delta^{-2}}-1$. The formulas for $\langle\chi_\epsilon^2\rangle$ and $\langle\phi_\epsilon^2\rangle$ were derived in Refs. 5 and 6 for the case of sound propagation in a turbulent atmosphere where $\sigma_\epsilon^2=\sigma_T^2/T_0^2$. The formulas for $\langle\chi_v^2\rangle$ and $\langle\phi_v^2\rangle$ are new results.

The second limiting case is $D\ll1$. This case corresponds to the geometric-acoustic approximation.[4] We shall also assume that $Dr^2/l^2\ll1$. All functions in the integrands of Eqs. (6), (7), (9), and (10) are expended into Taylor series over the small parameters D and Dr^2/l^2. Retaining only the first terms in these series yields the asymptotics of the correlation functions for log-amplitude and phase fluctuations

$$\begin{aligned}B_\chi(r)&=3\int_0^1 dt\ t^2(1-t)^2B_\chi^{pl}(rt)\\&=\frac{2\sqrt{\pi}x^3}{l^3}\int_0^1 dt\ t^2(1-t)^2e^{-r^2t^2/l^2}\\&\quad\times\left[\left(1-\frac{2r^2t^2}{l^2}+\frac{r^4t^4}{2l^4}\right)\sigma_\epsilon^2+\left(1-\frac{3r^2t^2}{l^2}+\frac{3r^4t^4}{2l^4}-\frac{r^6t^6}{6l^6}\right)\frac{12\sigma_v^2}{c_0^2}\right],\end{aligned}\tag{15}$$

$$\begin{aligned}B_\phi(r)&=\int_0^1 dt\ B_\phi^{pl}(rt)\\&=\frac{\pi k^2l^2x}{8r}\left[\operatorname{erf}\left(\frac{r}{l}\right)\sigma_\epsilon^2+\left(\frac{1}{2}\operatorname{erf}\left(\frac{r}{l}\right)+\frac{r}{l\sqrt{\pi}}\,e^{-r^2/l^2}\right)\frac{4\sigma_v^2}{c_0^2}\right].\end{aligned}\tag{16}$$

Here, $\operatorname{erf}(r/l)$ is the error function, B_χ^{pl} and B_ϕ^{pl} are the correlation functions for the log-amplitude and phase fluctuations of a plane wave, calculated in Paper I for the considered case $D\ll1$ and $Dr^2/l^2\ll1$. (The statistical moments of a plane wave have sub or superscripts ''pl'' in the present paper in order to distinguish them from analogous statistical moments of a spherical wave. However, they do not have such sub or superscripts in Paper I.) Equations (15) and (16) give us useful integral relationships between the correlation functions B_χ and B_χ^{pl}, and the correlation functions B_ϕ and B_ϕ^{pl}. The integral over t in Eq. (15) can be calculated analytically. The result is not, however, presented here because of its complexity.

The third limiting case $D\gg1$ corresponds to the Fraunhofer diffraction. We additionally suppose that $r^2/(l^2\ln D)\ll1$. If these inequalities are valid, the second terms in the braces in Eqs. (6), (7), (9), and (10) can be omitted. Then calculating the integrals over t in these equations yields

$$\begin{aligned}B_\chi(r)=B_\phi(r)&=\int_0^1 dt\ B_\chi^{pl}(rt)\\&=\int_0^1 dt\ B_\phi^{pl}(rt)\\&=\frac{\pi k^2l^2x}{16r}\left[\operatorname{erf}\left(\frac{r}{l}\right)\sigma_\epsilon^2+\left(\frac{1}{2}\operatorname{erf}\left(\frac{r}{l}\right)+\frac{r}{l\sqrt{\pi}}\,e^{-r^2/l^2}\right)\frac{4\sigma_v^2}{c_0^2}\right].\end{aligned}\tag{17}$$

Here, B_χ^{pl} and B_ϕ^{pl} are the correlation functions for the log-amplitude and phase fluctuations of a plane wave for the case $D\gg1$, calculated in Paper I. Equation (17) shows us a useful relationship between $B_{\chi,\phi}(r)$ and $B_{\chi,\phi}^{pl}(r)$. Using Eqs. (6),

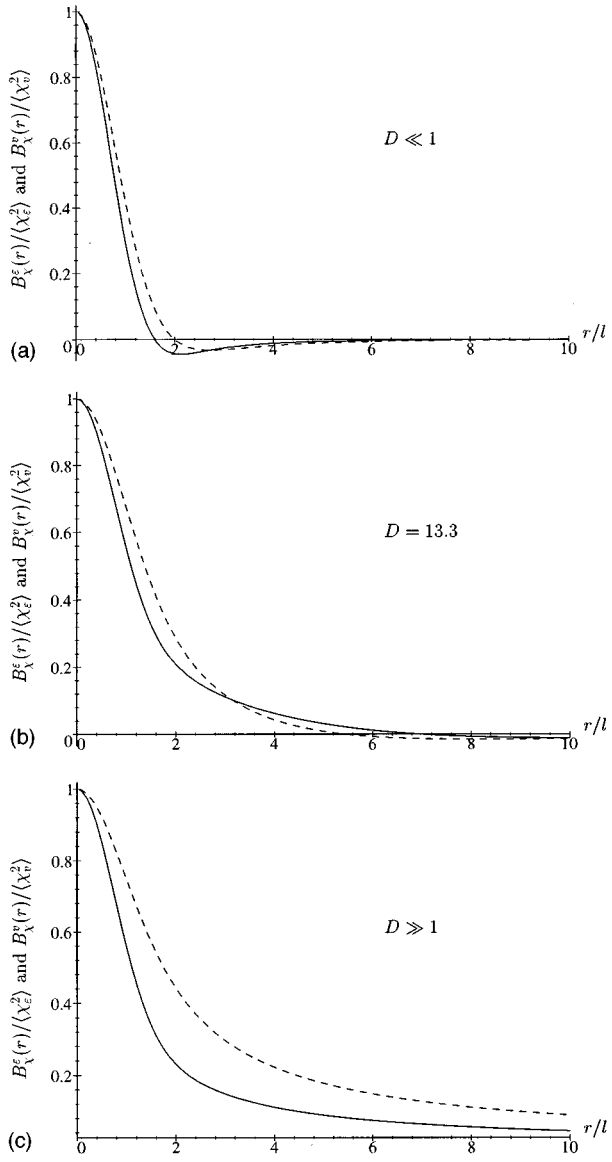


FIG. 1. The normalized temperature and medium velocity contributions to the correlation function for log-amplitude fluctuations, $B_\chi^\epsilon(r)/\langle\chi_\epsilon^2\rangle$ (dashed line) and $B_\chi^v(r)/\langle\chi_v^2\rangle$ (solid line), versus r/l for: $D\ll 1$ (a); $D=13.3$ (b); $D\gg 1$ (c).

(7), (14), (15), and (17), the normalized temperature and medium velocity contributions to the correlation function of log-amplitude fluctuations, $B_\chi^\epsilon(r)/\langle\chi_\epsilon^2\rangle$ and $B_\chi^v(r)/\langle\chi_v^2\rangle$, are plotted in Fig. 1 for different values of the wave parameter D: $D\ll 1$, $D=13.3$, and $D\gg 1$. It follows from the figure that the large the value of D, the greater the difference between these normalized contributions. In Fig. 1(a), the normalized temperature and medium velocity contributions to B_χ are nearly the same. Nevertheless, $B_\chi^\epsilon(r)$ and $B_\chi^v(r)$ differ significantly since $\langle\chi_v^2\rangle=3\langle\chi_\epsilon^2\rangle$ for $D\ll 1$.

Using Eqs. (9), (10), (12), (16), and (17), the normalized temperature and medium velocity contributions to the correlation function of phase fluctuations, $B_\phi^\epsilon(r)/\langle\phi_\epsilon^2\rangle$ and $B_\phi^v(r)/\langle\chi_v^2\rangle$, are plotted in Fig. 2 for $D=13.3$. The difference between these normalized contributions is more pronounced than that for the comparable case of log-amplitude fluctuations; see Fig. 1(b). It can be shown from Eqs. (12),

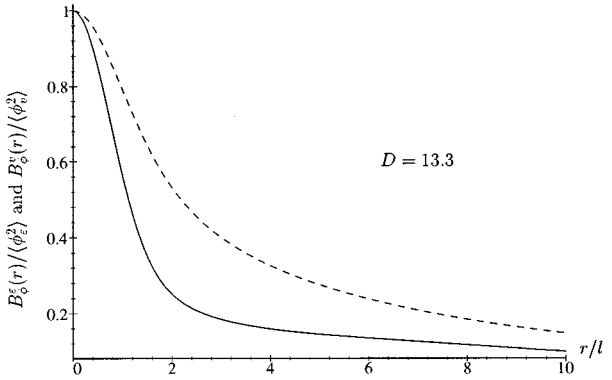


FIG. 2. The normalized temperature and medium velocity contributions to the correlation function for phase fluctuations, $B_\phi^\epsilon(r)/\langle\phi_\epsilon^2\rangle$ (dashed line) and $B_\phi^v(r)/\langle\phi_v^2\rangle$ (solid line), versus r/l for $D=13.3$.

(16), and (17) that the functions $B_\phi^\epsilon(r)/\langle\phi_\epsilon^2\rangle$, calculated for $D\ll 1$ and $D\gg 1$, are identical. Furthermore, using these equations and Eq. (11), one reveals that both functions coincide with $B_\chi^\epsilon(r)/\langle\chi_\epsilon^2\rangle$, calculated for the case $D\gg 1$ and plotted in Fig. 1(c). Similarly, using Eqs. (11), (12), (16), and (17), one reveals that the functions $B_\phi^v(r)/\langle\phi_v^2\rangle$, calculated for $D\ll 1$ and $D\gg 1$, are identical and coincide with $B_\chi^v(r)/\langle\chi_v^2\rangle$, calculated for $D\gg 1$ and shown in Fig. 1(c).

The correlation radii r_χ^ϵ, r_χ^v, r_ϕ^ϵ, and r_ϕ^v are important characteristics of the correlation functions $B_\chi^\epsilon(r)$, $B_\chi^v(r)$, $B_\phi^\epsilon(r)$, and $B_\phi^v(r)$, respectively, and are determined by equations similar to Eq. (20) from Paper I. Equations (6), (7), (9), and (10) allow one to compute these correlation radii. The

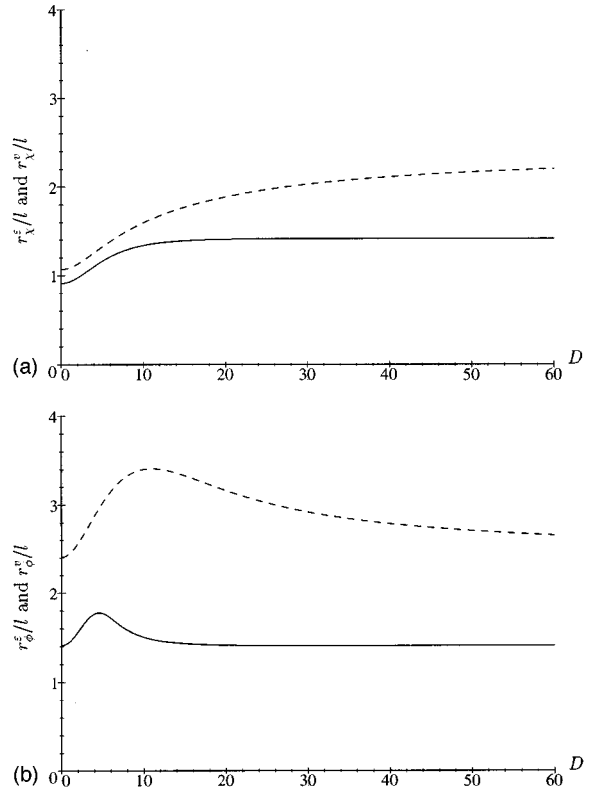


FIG. 3. The normalized correlation radii of the correlation functions for log-amplitude and phase fluctuations versus D: r_χ^ϵ/l (dashed line) and r_χ^v/l (solid line) (a); r_ϕ^ϵ/l (dashed line) and r_ϕ^v/l (solid line) (b).

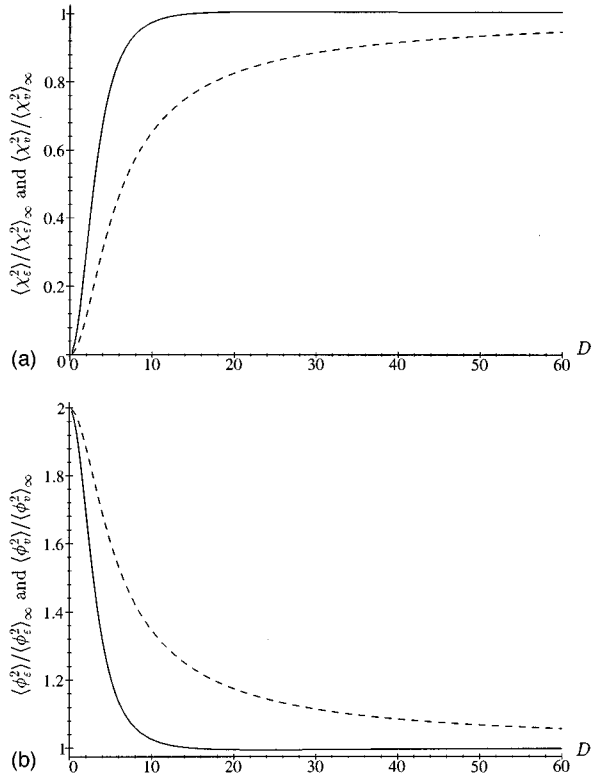


FIG. 4. The normalized temperature and medium velocity contributions to the variances of log-amplitude and phase fluctuations versus D: $\langle\chi_\epsilon^2\rangle/\langle\chi_\epsilon^2\rangle_\infty$ (dashed line) and $\langle\chi_v^2\rangle/\langle\chi_v^2\rangle_\infty$ (solid line) (a); $\langle\phi_\epsilon^2\rangle/\langle\phi_\epsilon^2\rangle_\infty$ (dashed line) and $\langle\phi_v^2\rangle/\langle\phi_v^2\rangle_\infty$ (solid line) (b).

normalized correlation radii for log-amplitude fluctuations r_χ^ϵ/l and r_χ^v/l vs D are plotted in Fig. 3(a), while the normalized correlation radii for phase fluctuations r_ϕ^ϵ/l and r_ϕ^v/l vs D are plotted in Fig. 3(b). It follows from these figures that the correlation radii r_χ^ϵ and r_ϕ^ϵ, associated with temperature fluctuations, are noticeably greater than their corresponding values r_χ^v and r_ϕ^v, associated with medium velocity fluctuations. The maximum of the ratio r_χ^ϵ/r_χ^v is 1.71 for $D\to\infty$, and the maximum of the ratio r_ϕ^ϵ/r_ϕ^v is 2.34 for $D=13.3$. The latter ratio also reaches the value 1.71 if $D\to\infty$.

C. Variance of log-amplitude and phase fluctuations

The variances of log-amplitude and phase fluctuations are given by Eqs. (11) and (12). It is worthwhile considering here the normalized temperature and medium velocity contributions to $\langle\chi^2\rangle$ and $\langle\phi^2\rangle$

$$\frac{\langle\chi_\epsilon^2\rangle}{\langle\chi_\epsilon^2\rangle_\infty}=1-F_1(D),\quad \frac{\langle\chi_v^2\rangle}{\langle\chi_v^2\rangle_\infty}=1-F_2(D),$$
$$\frac{\langle\phi_\epsilon^2\rangle}{\langle\phi_\epsilon^2\rangle_\infty}=1+F_1(D),\quad \frac{\langle\phi_v^2\rangle}{\langle\phi_v^2\rangle_\infty}=1+F_2(D). \tag{18}$$

Here, $\langle\chi_\epsilon^2\rangle_\infty=\langle\phi_\epsilon^2\rangle_\infty=\sqrt{\pi}k^2lx\sigma_\epsilon^2/8$ and $\langle\chi_v^2\rangle_\infty=\langle\phi_v^2\rangle_\infty=\sqrt{\pi}k^2lx\sigma_v^2/2c_0^2$ are the values of the variances for $D\to\infty$. These values coincide with those calculated in Paper I for plane wave propagation. The normalized temperature and medium velocity contributions to $\langle\chi^2\rangle$ and $\langle\phi^2\rangle$ vs D are plotted in Fig. 4(a) and (b). It follows from Fig. 4(a) that $\langle\chi_v^2\rangle/\langle\chi_v^2\rangle_\infty$ is greater than $\langle\chi_\epsilon^2\rangle/\langle\chi_\epsilon^2\rangle_\infty$ for all values of D. The maximum of the ratio $(\langle\chi_v^2\rangle/\langle\chi_v^2\rangle_\infty)/(\langle\chi_\epsilon^2\rangle/\langle\chi_\epsilon^2\rangle_\infty)$ is 3 for $D=0$. This ratio tends to 1 as $D\to\infty$. On the other hand, it follows from Fig. 4(b) that $\langle\phi_v^2\rangle/\langle\phi_v^2\rangle_\infty$ is always less than $\langle\phi_\epsilon^2\rangle/\langle\phi_\epsilon^2\rangle_\infty$. The minimum of the ratio $(\langle\phi_v^2\rangle/\langle\phi_v^2\rangle_\infty)/(\langle\phi_\epsilon^2\rangle/\langle\phi_\epsilon^2\rangle_\infty)$ is 0.74 for $D=6.68$. This ratio also tends to 1 as $D\to\infty$. Since $D=4x/(kl^2)$, it follows from Fig. 4 and Eqs. (11) and (12) that $\langle\chi_\epsilon^2\rangle$ and $\langle\chi_v^2\rangle$, and $\langle\phi_\epsilon^2\rangle$ and $\langle\phi_v^2\rangle$ have different dependencies on x, k, and l.

D. Coherence function of the sound field

Now consider the coherence function Γ of a spherical wave in a moving random medium. Substituting B_χ and B_ϕ from Eqs. (5)–(10) into Eq. (4) yields

$$\Gamma(\mathbf{r},\mathbf{r}_+)=\frac{1}{(4\pi x)^2}\exp\left\{\frac{ik\mathbf{r}\cdot\mathbf{r}_+}{x}-\frac{\sqrt{\pi}k^2lx}{4}\right.$$
$$\times\int_0^1 dt\left[(1-e^{-r^2t^2/l^2})\sigma_\epsilon^2\right.$$
$$\left.\left.+\left(1-\left(1-\frac{r^2t^2}{l^2}\right)e^{-r^2t^2/l^2}\right)\frac{4\sigma_v^2}{c_0^2}\right]\right\}. \tag{19}$$

Calculating the integral over t in this formula, one obtains

$$\Gamma(\mathbf{r},\mathbf{r}_+)=\frac{e^{ik\mathbf{r}\cdot\mathbf{r}_+/x}}{(4\pi x)^2}\,\Gamma_\epsilon(r)\Gamma_v(r). \tag{20}$$

Here, Γ_ϵ and Γ_v are the factors of Γ due to sound scattering by temperature and medium velocity fluctuations, given by

$$\Gamma_\epsilon(r)=\exp\left\{-2\gamma_\epsilon x\left[1-\frac{\sqrt{\pi}l}{2r}\,\mathrm{erf}\left(\frac{r}{l}\right)\right]\right\}, \tag{21}$$

$$\Gamma_v(r)=\exp\left\{-2\gamma_v x\left[1-\frac{\sqrt{\pi}l}{4r}\,\mathrm{erf}\left(\frac{r}{l}\right)-\frac{1}{2}e^{-r^2/l^2}\right]\right\}, \tag{22}$$

where γ_ϵ and γ_v are the temperature and medium velocity contributions to the attenuation coefficient γ, which are determined by Eq. (22) from Paper 1. Equation (21) for Γ_ϵ is known in the literature,[7] while Eq. (22) is a new result. It follows from these equations that $\Gamma_\epsilon\neq\Gamma_v$, i.e., the temperature and medium velocity contributions to the coherence function are different.

Let r_Γ^ϵ and r_Γ^v be the coherence radii of the sound field in media with temperature and medium velocity fluctuations, respectively, defined by the following equations: $\Gamma_\epsilon(r_\Gamma^\epsilon)=\Gamma_\epsilon(0)/e$ and $\Gamma_v(r_\Gamma^v)=\Gamma_v(0)/e$. It follows from Eqs. (21) and (22) that r_Γ^ϵ and r_Γ^v satisfy the equations:

$$2\gamma_\epsilon x\left[1-\frac{\sqrt{\pi}l}{2r_\Gamma^\epsilon}\,\mathrm{erf}\left(\frac{r_\Gamma^\epsilon}{l}\right)\right]=1, \tag{23}$$

$$2\gamma_v x\left[1-\frac{\sqrt{\pi}l}{4r_\Gamma^v}\,\mathrm{erf}\left(\frac{r_\Gamma^v}{l}\right)-\frac{1}{2}e^{-(r_\Gamma^v/l)^2}\right]=1. \tag{24}$$

It can be shown from Eqs. (23) and (24) that r_Γ^ϵ and r_Γ^v exist only if $\gamma_\epsilon x>1/2$ and $\gamma_v x>1/2$, respectively. This is ex-

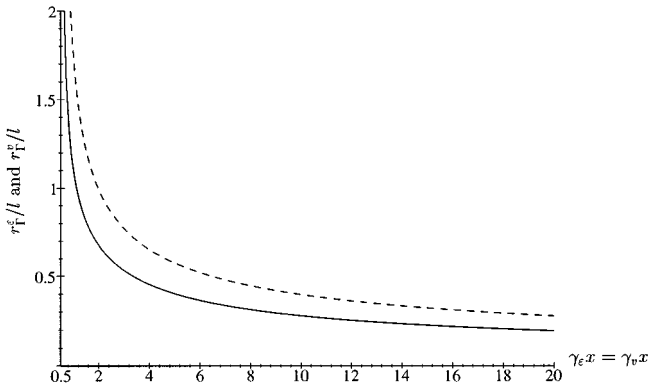


FIG. 5. The normalized coherence radii r_Γ^ϵ/l (dashed line) and r_Γ^v/l (solid line) versus $\gamma_\epsilon x = \gamma_v x$.

plained by the fact that $\Gamma_\epsilon(r) > \Gamma_\epsilon(0)/e$ and $\Gamma_v(r) > \Gamma_v(0)/e$ for $\gamma_\epsilon x < 1/2$ and $\gamma_v x < 1/2$. Furthermore, it can be shown from Eqs. (23) and (24) that $r_\Gamma^\epsilon \to \infty$ and $r_\Gamma^v \to \infty$ if $\gamma_\epsilon x \to 1/2$ and $\gamma_v x \to 1/2$. Also it follows from these equations that $r_\Gamma^\epsilon = \sqrt{3/(2\gamma_\epsilon x)}\, l = \sqrt{3} r_{\Gamma,pl}^\epsilon$ and $r_\Gamma^v = \sqrt{3/(4\gamma_v x)}\, l = \sqrt{3} r_{\Gamma,pl}^v$ if $\gamma_\epsilon x \gg 1$ and $\gamma_v x \gg 1$, respectively. Here, $r_{\Gamma,pl}^\epsilon$ and $r_{\Gamma,pl}^v$ are the coherence radii of a plane wave, obtained in Paper I.

The coherence radii r_Γ^ϵ and r_Γ^v, normalized to l, are plotted in Fig. 5 as functions of $\gamma_\epsilon x = \gamma_v x$. It follows from the figure that r_Γ^ϵ is greater than r_Γ^v for all values of $\gamma_{\epsilon,v}\, x$.

II. COMPARISON WITH EXPERIMENTAL DATA

An extensive experimental study of line-of-sight sound propagation in a turbulent atmosphere is reported in Ref. 7. The solid points in Fig. 6 are the measured values of the variance of phase fluctuations $\langle \phi^2 \rangle$ versus the sound frequency $f = kc/2\pi$, taken from Fig. 3 in Ref. 7. The short dashed line 1 represents the theoretical dependence of the variance of phase fluctuations

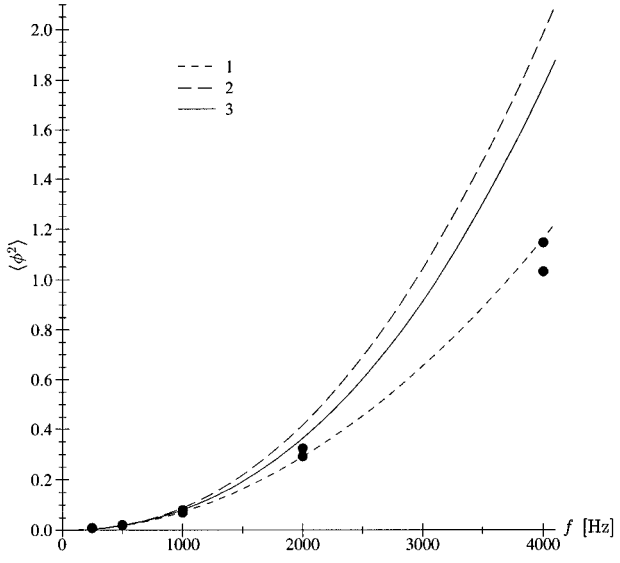


FIG. 6. The variance of phase fluctuations $\langle \phi^2 \rangle$ versus the frequency f. Solid points are measured values of $\langle \phi^2 \rangle$. The theoretical predictions of $\langle \phi^2 \rangle$ on f are given by the lines 1 [based on Eq. (25)], 2 [based on Eq. (26)], and 3 [based on Eq. (12)].

$$\langle \phi^2 \rangle_\infty = \frac{\sqrt{\pi} k^2 l x}{8} \left(\frac{\sigma_T^2}{T_0^2} + \frac{4\sigma_v^2}{c_0^2} \right) \tag{25}$$

on frequency f for the limiting case $D \to \infty$. Equation (25) can be derived from Eq. (12). Indeed, in the latter equation, $\sigma_\epsilon^2 = \sigma_T^2/T_0^2$ for the considered case of sound propagation in the atmosphere. Furthermore, it can be shown from Eqs. (13) and (14) that $F_1(D=\infty) = F_2(D=\infty) = 0$. Substitution of these equalities into Eq. (12) yields Eq. (25). When plotting the line 1, the values of $l = 1.1$ m, $x = 85$ m, and $\langle \mu^2 \rangle = \sigma_T^2/4T_0^2 + \sigma_v^2/c_0^2 = 2.6 \times 10^{-6}$ are used, which were measured during the experiment.[7,8] It can be shown that Eq. (25) and line 1 in Fig. 6 coincide with Eq. (2) and the theoretical curve in Fig. 3 from Ref. 7, respectively.

It follows from Fig. 6 that experimental data from Ref. 7 are in good agreement with the variance $\langle \phi^2 \rangle_\infty$ calculated for the limiting case $D \to \infty$. However, this limiting case should not be used for comparing with the experimental data. Indeed, for the values of l and x, given above, the wave parameter D is not very large and is equal to 3.8 and 7.7 for $f = 4$ kHz and 2 kHz, respectively.

Therefore it is worthwhile to compare the experimental data, presented in Fig. 6, with theoretical predictions for $\langle \phi^2 \rangle$, which are valid for arbitrary values of D. In the previous theories of sound propagation in the turbulent atmosphere, $\langle \phi^2 \rangle$ was given by [e.g., see Eqs. (7)–(10) in Ref. 6]

$$\langle \phi^2 \rangle = \frac{\sqrt{\pi} k^2 l x}{8} (1 + F_1(D)) \left(\frac{\sigma_T^2}{T_0^2} + \frac{4\sigma_v^2}{c_0^2} \right). \tag{26}$$

The dashed line 2 in Fig. 6 is the theoretical dependence of $\langle \phi^2 \rangle$ on f, plotted in accordance with Eq. (26). As seen in figure this line is located above the experimental points for $f \geqslant 2$ kHz.

Now let us compare the same experimental data with the theoretical predictions of the present paper, i.e., with Eq. (12). Since Ref. 7 only provided the value of $\langle \mu^2 \rangle$, we used $\sigma_v/\sigma_T = 0.6(\mathrm{m/s})/^\circ\mathrm{C}$ as suggested by Daigle as a best estimate for the day when the experimental data involved were measured (a hot and sunny day with light wind). For given values of $\langle \mu^2 \rangle$ and σ_v/σ_T, it is trivial to calculate the values of σ_v^2 and σ_T^2 entering into Eq. (12). The solid line 3 in Fig. 6 is the dependence of $\langle \phi^2 \rangle$ on f calculated by using this equation. This line fits the experimental data somewhat better than line 2, based on the previous theories.

The measured variance of log-amplitude fluctuations $\langle \chi^2 \rangle$ is also reported in Ref. 7. As indicated by data in Ref. 7, the measured values of $\langle \chi^2 \rangle$ are significantly smaller than those of $\langle \phi^2 \rangle$. This result cannot be explained by the previous theories nor the theory developed in the present paper, which predict that $\langle \chi^2 \rangle \sim \langle \phi^2 \rangle$ for large enough values of D. A possible explanation of this result is that the Gaussian spectrum is not appropriate for theoretical predictions of $\langle \chi^2 \rangle$, which is affected by small-scale inhomogeneities of the order of the Fresnel zone $\sqrt{\lambda x}$. On the other hand, phase fluctuations are affected by much larger inhomogeneities, a spectrum of which can be approximated by the Gaussian spectrum.

III. CONCLUSIONS

In the present paper, line-of-sight propagation of a spherical sound wave in a moving random medium has been considered. For line-of-sight sound propagation, the statistical moments of a sound field, which are most often of interest, are the following: the variances of log-amplitude and phase fluctuations, $\langle \chi^2 \rangle$ and $\langle \phi^2 \rangle$, the correlation functions of log-amplitude and phase fluctuations, $B_\chi(r)$ and $B_\phi(r)$, the mean sound field $\langle p \rangle$, and the coherence function $\Gamma(\mathbf{r},\mathbf{r}_+)$ of the sound field. The formulas for these statistical moments have been obtained for arbitrary (anisotropic) spectra of temperature and medium velocity fluctuations. Then, these statistical moments have been calculated analytically and numerically for the Gaussian spectra of temperature and medium velocity fluctuations. The temperature and medium velocity contributions to these statistical moments have been compared. It has been shown that these contributions can differ not only quantitatively but also qualitatively, i.e., they can have different dependencies on x, k, and l.

The paper also presents useful relationships between the statistical moments of a spherical wave propagating in a moving random medium and analogous statistical moments of a plane wave, obtained in Paper I.

The variance of phase fluctuations $\langle \phi^2 \rangle$, calculated in the paper, is compared with experimental results presented in Ref. 7. It is shown that the obtained theoretical results are in somewhat better agreement with the experimental data than those based on the previous theories. On the other hand, the variance of log-amplitude fluctuations $\langle \chi^2 \rangle$ calculated in the present paper and that calculated by using the previous theories do not agree with the experimental results.

ACKNOWLEDGMENTS

This material is based upon work supported in part by the U.S. Army Research Office under Contract No. DAAH04-95-1-0593. We would like to thank Dr. G. Daigle for his help in comparing the theoretical results obtained with experimental data presented in Ref. 7. Also we would like to thank the reviewers and the Associate Editor Louis Sutherland for many useful comments to the present paper and Paper I.

APPENDIX: CALCULATIONS OF THE CORRELATION FUNCTIONS OF LOG-AMPLITUDE AND PHASE FLUCTUATIONS

In the Appendix, Eqs. (5)–(10) for the correlation functions $B_\chi(r)$ and $B_\phi(r)$ of log-amplitude and phase fluctuations are derived. Replacing $\Phi_{\text{eff}}(0,K)$ in Eq. (1) by its value given by Eq. (8) from Paper I yields

$$B_{\chi,\phi}(r) = \frac{\sqrt{\pi} k^2 l x}{4} \int_0^1 dt \int_0^\infty dK\ K[1 \mp \cos(K^2 D\tau)] \times e^{-K^2} J_0(2K\rho) \left(\sigma_\epsilon^2 + \frac{4\sigma_v^2 K^2}{c_0^2} \right), \tag{A1}$$

where $\rho = rt/l$, $\tau = t(1-t)$, and the upper and lower signs on the right-hand side of this equation correspond to B_χ and B_ϕ, respectively. Equation (A1) can be expressed in the form

$$B_{\chi,\phi}(r) = \frac{\sqrt{\pi} k^2 l x}{4} \int_0^1 dt \left[(I_1 \mp I_2)\sigma_\epsilon^2 + (I_3 \mp I_4) \frac{4\sigma_v^2}{c_0^2} \right], \tag{A2}$$

where

$$I_1 = \int_0^\infty dK\ K J_0(2K\rho) e^{-K^2}, \tag{A3}$$

$$I_2 = \Re \int_0^\infty dK\ K J_0(2K\rho) e^{-(1+iD\tau)K^2}, \tag{A4}$$

$$I_3 = \int_0^\infty dK\ K^3 J_0(2K\rho) e^{-K^2}, \tag{A5}$$

$$I_4 = \Re \int_0^\infty dK\ K^3 J_0(2K\rho) e^{-(1+iD\tau)K^2}. \tag{A6}$$

Integrals I_{1-4} can be calculated by making use of the reference integral[9]

$$I = \int_0^\infty dK\ K^{\alpha-1} J_0(2K\rho) e^{-qK^2} = \frac{\Gamma(\alpha/2)}{2q^{\alpha/2}} {}_1F_1(\alpha/2;1;-\rho^2/q), \tag{A7}$$

where $\Re\alpha > 0$. This integral is similar to the reference integral (A7) in Paper I. I_1 becomes I if $\alpha = 2$ and $q = 1$. Taking also into account Eq. (A8) from Paper I yields

$$I_1 = (1/2)\exp(-\rho^2). \tag{A8}$$

I_2 becomes I if $\alpha = 2$ and $q = 1 + iD\tau$. Using Eq. (A8) from Paper I, one obtains

$$I_2 = \frac{1}{2} \Re \left(\frac{e^{-\rho^2/(1+iD\tau)}}{1+iD\tau} \right) = \frac{e^{-\rho^2/d}}{2d} \left[\cos\left(\frac{\rho^2 D\tau}{d} \right) + D\tau \sin\left(\frac{\rho^2 D\tau}{d} \right) \right]. \tag{A9}$$

Here, $d = 1 + D^2\tau^2$. I_3 becomes I if $\alpha = 4$ and $q = 1$. Taking also into account that

$${}_1F_1(2;1;y) = (1+y)\exp(y), \tag{A10}$$

yields

$$I_3 = (1/2)(1-\rho^2)\exp(-\rho^2). \tag{A11}$$

Finally, I_4 becomes I if $\alpha = 4$ and $q = 1 + iD\tau$. Using Eq. (A11), one obtains

$$I_4 = \frac{1}{2} \Re \left(\frac{1-\rho^2/(1+iD\tau)}{(1+iD\tau)^2} e^{-\rho^2/(1+iD\tau)} \right) = \frac{e^{-\rho^2/d}}{2d^3} \left[(1 - D^4\tau^4 + \rho^2(3d-4)) \cos\left(\frac{\rho^2 D\tau}{d} \right) + D\tau(2d + \rho^2(d-4)) \sin\left(\frac{\rho^2 D\tau}{d} \right) \right]. \tag{A12}$$

Substituting I_1, I_2, I_3, and I_4 into Eq. (A2), one obtains formulas for B_χ and B_ϕ, which coincide with Eqs. (5)–(7) and Eqs. (8)–(10), respectively.

[1] V. E. Ostashev, V. Mellert, R. Waudelt, and F. Gerdes, ''Propagation of sound in a turbulent medium. I. Plane waves,'' J. Acoust. Soc. Am. **102**, 2561–2570 (1997).

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

[3] V. E. Ostashev, ''Sound propagation and scattering in media with random inhomogeneities of sound speed, density and medium velocity (Review Article),'' Waves Random Media **4**, 403–428 (1994).

[4] S. M. Rytov, Yu. A. Kravtzov, and V. I. Tatarskii, *Principles of Statistical Radio Physics. Part 4. Wave Propagation through Random Media* (Springer Verlag, Berlin, 1989).

[5] V. N. Karavainikov, ''Fluctuations of amplitude and phase in a spherical wave,'' Sov. Phys. Acoust. **3**, 175–186 (1956).

[6] G. A. Daigle, J. E. Piercy, and T. F. W. Embleton, ''Effects of atmospheric turbulence on the interference of sound waves near a hard boundary,'' J. Acoust. Soc. Am. **64**, 622–630 (1978).

[7] G. A. Daigle, J. E. Piercy, and T. F. W. Embleton, ''Line-of-sight propagation through atmospheric turbulence near the ground,'' J. Acoust. Soc. Am. **74**, 1505–1513 (1983).

[8] Note that to an accuracy considered in Ref. 7, the value of $\langle \mu^2 \rangle = 2.6 \times 10^{-6}$ is in agreement with 3×10^{-6} presented in Fig. 3 in Ref. 7. Only for the former value of $\langle \mu^2 \rangle$, the theoretical prediction of $\langle \phi^2 \rangle_\infty$, based on Eq. (25) or Eq. (2) in Ref. 7, coincides with the theoretical curve in this figure.

[9] A. P. Prudnikov, Y. A. Brychkov, and O. I. Marichev, *Integrals and Series (Elementary Functions)* (Nauka, Moscow, 1981) (in Russian).

Theory of acoustic attenuation, dispersion, and pulse propagation in unconsolidated granular materials including marine sediments

Michael J. Buckingham[a)]
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0213

(Received 7 August 1996; revised 25 November 1996; accepted 31 July 1997)

A unified theory of sound propagation in saturated marine sediments is developed on the basis of a linear wave equation, which includes a new dissipation term representing internal losses arising from interparticle contacts. This loss mechanism, which shows a ''memory'' or hysteresis, is proposed as being responsible for the acoustic properties of sediments. To accommodate the memory, the loss term in the wave equation is formulated as a temporal convolution between the particle velocity and a material response function, $h(t)$, which varies as t^{-n}, where $0<n<1$. The compressional wave that emerges from the analysis shows: (1) an attenuation that scales almost exactly as the first power of frequency (corresponding to a constant Q) over an unlimited number of decades; and (2) weak logarithmic dispersion. As well as being characteristic of the wave properties of actual marine sediments, the predicted attenuation and dispersion are consistent with the Kronig–Kramers relationships. The theory also leads to pulse propagation that is strictly causal, which, although a necessary requirement, is an issue that has been widely debated in the literature in the context of an attenuation that is proportional to frequency. Finally, the wave properties (phase speed and attenuation) are related to the mechanical properties (grain size, density, and porosity) of a sediment by combining the Hertz theory of particles in contact with a new, rough-surface, random-packing model of mineral grains in unconsolidated granular media. The resultant relationships between the acoustical and mechanical properties (e.g., sound speed and porosity) of marine sediments are shown to follow the trends of published experimental data sets very closely.
 [S0001-4966(97)04411-1]

PACS numbers: 43.30.Ma [JHM]

LIST OF SYMBOLS

c_p	compressional wave speed (m/s)
c_0	compressional wave speed in absence of intergranular friction (m/s)
α_p	compressional attenuation coefficient (nepers/m)
β_p	compressional loss tangent
Q	quality factor
χ_f	compressional dissipation coefficient
μ_c	compressional frictional rigidity modulus
ω	angular frequency
k_0	wave number in absence of intergranular friction
t	time
p	pressure fluctuation
v	particle velocity
ρ	density fluctuation
$\psi(t)$	velocity potential
$\Psi(j\omega)$	Fourier transform of $\psi(t)$
$h(t)$	material memory function
$H(j\omega)$	Fourier transform of $h(t)$
n	material memory exponent $(0<n<1)$
u_g	mean grain diameter, microns
N	porosity $(0<N<1)$
ρ_0	bulk density of sediment (kg/m^3)
κ	bulk modulus of sediment (Pa)
γ	volume ratio of smaller to larger grains in a bimodal sediment
$\rho_w=1024$ kg/m^3	density of pore water
$\rho_g=2700$ kg/m^3	density of mineral grains
$\kappa_w=2.25\times10^9$ Pa	bulk modulus of pore water
$\kappa_g=1.47\times10^{10}$ Pa	bulk modulus of mineral grains
$\Delta=3$ μm	rms particle roughness
$P=0.63$	packing factor of randomly packed smooth spheres
$u_0=1000$ μm	reference grain diameter
$\mu_0=2\times10^9$ Pa	compressional frictional rigidity constant

INTRODUCTION

A substantial body of data supports the view that compressional-wave attenuation in many porous, granular materials varies more or less accurately as the first power of frequency, f^1, over an extended frequency range, from 1 Hz up to 1 MHz. The evidence for this observation has been extensively discussed in a nicely argued review of the sub-

[a)]Also affiliated to: Institute of Sound and Vibration Research The University, Southampton SO17 1BJ, England.

ject by Kibblewhite,[1] which includes a comprehensive bibliography covering much of the relevant literature.

A striking conclusion from Kibblewhite's[1] careful examination of the available data is that, in at least one class of granular materials, i.e., dry sandstones, the attenuation scales accurately with the first power of frequency from seismic to high ultrasonic frequencies, a frequency range of some six decades. The fact that in dry, consolidated, porous media the attenuation varies as f^1 over so many decades suggests that this type of attenuation may be attributed to a single loss mechanism, which can be identified from the geophysics literature as an unspecified form of internal dissipation[2–5] arising at grain-to-grain contacts.

The question of whether attenuation in saturated, unconsolidated marine sediments is accurately proportional to the first power of frequency is still under debate. Hamilton[6,7] has long argued, on the basis of extensive experimental evidence, that attenuation in marine sediments does indeed exhibit an f^1 dependence, a point of view which is shared by other investigators.[8] On the other hand, Kibblewhite[1] concludes from the data of Hamilton[6,7] and others that attenuation in fluid-saturated sediments does not accurately follow the f^1 law. In these materials, he argues, there appear to be two mechanisms responsible for the attenuation, one being ''internal friction,'' which somehow gives rise to a linear scaling with frequency, as in dry materials, and a second due to the viscosity of the pore fluid.[5] The effects of viscosity, if any, are said to manifest themselves globally through the relative motion of the pore fluid and the sediment frame (if this has any meaning in the context of an unconsolidated medium), and also locally through fluid motion in the neighbourhood of interparticle contacts. According to Kibblewhite's argument, these additional viscous loss components are responsible for any deviation from the f^1 law in saturated materials.

It is evident from Kibblewhite's discussion[1] that internal dissipation, which is frequently stated (without proof) as giving rise to the proportionality between attenuation and frequency, is a fundamentally important physical phenomenon exhibited by granular materials,[9] irrespective of whether they are dry or wet. In the case of the latter, it seems that viscous losses could also be present, but to a degree that is still uncertain. The term *characteristic attenuation* is introduced here to identify the component of attenuation that scales accurately as f^1, corresponding to a constant quality factor,[10] or Q. (A possible alternative term is *intrinsic attenuation*, but this is already in use[1] to describe all internal loss mechanisms, such as internal friction and pore-fluid viscosity, but excluding losses due to scattering and also energy conversion losses arising from coupling between compressional and shear waves.)

A number of theories of attenuation in granular materials have been explored over the years. An early mechanism, discussed by Morse,[11] is bulk viscosity, corresponding to a frictional stress that is proportional to the particle velocity. However, viscous dissipation yields an attenuation that scales as the square of frequency[12] at low frequencies and as the square root of frequency[13] beyond some critical frequency, and hence cannot account for the characteristic attenuation. Porosity is another candidate mechanism, but it is well-known that a simple porous solid with a rigid frame, satisfying Darcy's law,[14] gives rise to an attenuation that is constant at high frequencies and scales as the square root of frequency below a critical frequency.[13,15] Again, such behavior does not match that of the characteristic attenuation.

Certain theoretical implications of a linear scaling between attenuation and frequency have been discussed by Horton[16] on the basis of the Kronig–Kramers dispersion relationships. He initially concluded that there is no dispersion in a material showing the characteristic attenuation, which is consistent with the frequency-independent sound speed observed in several field measurements.[17,18] However, in a later paper, Horton,[19] acknowledging an algebraic error in his previous analysis, pointed out that when the attenuation scales with frequency the phase speed actually shows a very slight logarithmic dispersion, in agreement with the theoretical treatment of O'Donnell *et al.*[20]

Laboratory measurements of logarithmic dispersion in saturated medium sand have been reported by Wingham,[21] at a level that agrees with O'Donnell *et al.*'s theoretical expression. However, the dispersion is so weak that a sensitive differential technique was required to detect it, which may account for the fact that it was not observed in the field measurements of McDonal[17] and Hamilton.[18] White[13] has collated a number of hypothetical attenuation-dispersion pairs that have been proposed by various authors in efforts to model the characteristic attenuation in granular materials.

For a number of years, Hamilton[22,23] has taken an empirical approach to modeling the geoacoustic properties of marine sediments. From a practical point of view, he has had great success with his models, although he does not consider the specific mechanism responsible for the attenuation. However, he does allow for the possibility of both dispersion and an attenuation that deviates from an f^1 power law.

An entirely different modeling approach is found in the classical works of Biot on the mechanics of porous, deformable media.[24–26] In his original theory, Biot assumed Poiseuille flow in a saturated porous medium with a linear elastic frame. From this model, the functional dependence of the attenuation on frequency is the same at low frequencies as that due to viscosity, scaling with the second power of frequency, f^2; and at higher frequencies it varies as f^0 (a constant). After modifying the theory to include the effects of viscosity in the vicinity of intergranular contacts, the frequency dependence of the attenuation takes the same functional form as in the case of a viscous fluid: At low frequencies the attenuation scales with f^2 and at higher frequencies it is proportional to $f^{1/2}$. Thus the Biot theory, even in its modified form, does not account for the characteristic attenuation in granular materials.

Biot's theoretical developments have been extended by Stoll[27] in a long-standing effort to model the geoacoustic behavior of marine sediments. The Biot–Stoll theory is a ''pseudo-harmonic'' analysis in which the shear and bulk moduli of an assumed skeletal frame are treated as complex quantities and assigned a frequency dependence determined from a match to data. This hybrid approach, a combination of analysis and empiricism, has had considerable success in

providing an accurate description of the geoacoustic behavior of many materials of interest to the geophysics community. However, the theory does not easily account for the characteristic attenuation exhibited by granular materials over an extended frequency range. Moreover, issues of causality and pulse propagation in porous materials, which have to be handled in the time domain, are not readily addressed within the pseudo-harmonic framework of the Biot–Stoll theory.

Although some form of internal friction has been identified experimentally as being responsible for the characteristic attenuation exhibited by granular materials,[2–5] the actual mechanism has not yielded readily to theoretical analysis. Various types of internal loss have been investigated with limited success, including a nonlinear mechanism associated with the motion of dislocations,[28] and frictional dissipation arising from the sliding of crack surfaces in contact with each other.[29] A symbolic manipulation approach to the problem of the characteristic attenuation has been developed by Verweij,[30] based on theoretical techniques developed by himself[31] and de Hoop.[32,33]

I. A NEW THEORY OF SEDIMENTS ACOUSTICS

In this article, an internally consistent, unified theory of acoustic propagation in a saturated, unconsolidated, marine sediment is developed. It is taken as axiomatic that an unconsolidated sediment is a two-phase medium, consisting of mineral particles and seawater, but possessing no rigid frame; that is to say, the elastic rigidity modulus of the medium is identically zero. Thus the material is treated as a bulk fluid in which sound propagation is governed by internal losses arising at grain-to-grain contacts. Throughout the discussion, the word ''friction'' is used exclusively and specifically to mean a new type of intergranular dissipation, which has fundamentally different properties from more familiar loss mechanisms such as viscosity or Coulomb friction.

The analysis is based on a linear wave equation in which intergranular dissipation is represented by a loss term that takes account of the hysteresis, or memory, of granular media. The effect of the memory on wave propagation is accommodated by setting the frictional stress equal to a temporal convolution between the particle velocity and a material memory function, $h(t)$. This formulation is identically that for a viscous fluid when $h(t)$ is set equal to a Dirac delta function, corresponding to the special case of a medium with no memory. A marine sediment, however, does possess a memory, the appropriate memory function being of the form $h(t) \propto t^{-n}$, where $0<n<1$. The properties of such a medium are quite different from those of say a viscous fluid.

Several interesting results emerge from the theory, including an attenuation coefficient that is almost exactly proportional to frequency over an unlimited number of decades. The model also predicts weak logarithmic dispersion which has exactly the form required by the Kronig–Kramers relationships, as well as faithfully matching that observed experimentally by Wingham.[21] An important prediction of the theory is that, with increasing frictional dissipation, the material becomes stiffer and consequently the compressional wave speed increases. In the time domain, an expression is derived for the Green's function, representing transient propagation, which is shown to be strictly causal (as of course it must be). As the intergranular friction is allowed to go to zero, the Green's function reduces in a natural way to the familiar retarded potential solution for the pulse shape in a lossless medium.

Viscosity of the pore fluid has often been suggested as being a significant loss mechanism in sediments.[27] The effects of pore fluid viscosity are examined briefly by including a conventional viscous dissipation term in the new wave equation in addition to the intergranular friction term. A solution for the attenuation coefficient is obtained which deviates at higher frequencies from the f^1 law due to intergranular friction alone; and the viscosity also introduces a greatly enhanced dispersion. Such behavior is not consistent with Wingham's[21] measurements of attenuation and dispersion in saturated sand, nor with the absence of dispersion in marine sediments reported by Hamilton.[18] This lack of experimental support provides strong evidence that, compared with intergranular friction, pore-fluid viscosity is insignificant in saturated marine sediments.

It has been confirmed experimentally[18] that the compressional wave speed in an unconsolidated sediment correlates well with the mechanical parameters (particle size, density, and porosity) of the medium, and that the porosity and density themselves are each correlated with the particle size. To establish theoretical relationships between porosity, density, and particle size, a new rough-particle, random-packing model of mineral particles is introduced. From this model of mechanical structure of the sediment, the porosity and the density are expressed as algebraic functions of the particle size.

To provide the link between the acoustic and mechanical properties of a sediment, the intergranular friction is related to the particle size by a fractional power law, as suggested by the Hertz theory of elastic spheres in contact. The final result is a set of equations which give all the acoustic and mechanical properties of sediments as simple algebraic functions of the particle size. These theoretical expressions, relating sound speed, attenuation, porosity, and density to the particle size, are valid for surficial sediments. It turns out that the theoretical relationships predicted by the combined geoacoustic models show compelling agreement with extensive, published data sets representing not only silts and all grades of sand but also the very fine-grained clays.

Sands and silts conform to the geometry of the rough-sphere model, in that the grains resemble distorted spheres. Clay particles, however, are not spherical but show a platelet structure with a high aspect ratio, which raises the question as to why they should be so well described by the rough-sphere theory? It is argued that a mineral platelet in clay may be thought of as a special (extreme) case of a rough sphere in which the magnitude of the roughness is comparable to or greater than the size of the sphere itself. In this picture, a platelet is essentially all roughness on a relatively small spherical body. Although this may be an unusual geometrical description of a clay particle, it appears to be accurate as far as packing arguments are concerned, and it brings the clays

within the unified geoacoustic theory of sediments developed here.

II. A WAVE EQUATION WITH INTER-GRANULAR FRICTION

To address the question of wave propagation in granular materials, the medium is treated as fluidlike, macroscopically homogeneous, isotropic, time invariant, and infinite. A source is assumed to excite compressional plane waves in the material and, since there are no boundaries, there is no coupling into shear waves. The propagation of the compressional waves is governed by a wave equation in which intergranular friction is represented by a very specific dissipation term, whose formulation is crucial to the argument.

Since a form of internal friction is the loss mechanism, the dissipation term must originate in the equation of motion,[34] which, for one-dimensional propagation in the x direction, is postulated to be

$$\rho_0 \frac{\partial v}{\partial t} + \frac{\partial p}{\partial x} - \left(\frac{4}{3}\,\eta_f + \lambda_f \right) \frac{\partial^2 [h(t) \otimes v(t)]}{\partial x^2} = 0, \tag{1}$$

where v is particle velocity in the x direction, p is the pressure fluctuation, t is time, ρ_0 is the bulk density, and the symbol $\otimes$ denotes a temporal convolution operation. The dissipation coefficients η_f and λ_f have been introduced by analogy with the shear and bulk viscosities of a conventional fluid, although in this case the subscript f identifies these coefficients as being associated with intergranular friction rather than viscosity.

The proposed dissipation term in Eq. (1) possesses something of the character of viscosity, in that the differential operator in the loss term is $\partial^2/\partial x^2$, but the operand, instead of being particle velocity, is the convolution of particle velocity with an impulse response function of the material, $h(t)$. Clearly, when $h(t)$ is a Dirac delta function the convolution reduces to the particle velocity, in which case the equation of motion is identically that for a viscous fluid[12] or a Voigt solid.[13] The more general form in Eq. (1) allows for the possibility that the medium ''remembers'' previous states, with a memory that is characterized by $h(t)$. A generalized description of linear internal losses in terms of convolution operations is fairly well-established in the geophysics literature.[35]

The remaining expressions necessary to derive the wave equation are standard in that they do not involve dissipation terms. Conservation of mass requires that

$$\rho_0 \frac{\partial v}{\partial x} + \frac{\partial \rho}{\partial t} = 0, \tag{2}$$

and the equation of state is

$$p = c_0^2 \rho, \tag{3a}$$

where

$$c_0 = \sqrt{\frac{\kappa}{\rho_0}}. \tag{3b}$$

In these equations, ρ is the density fluctuation and κ is the bulk modulus of the medium. According to Eq. (3b), c_0 may be interpreted as the sound speed in the medium in the absence of grain-to-grain losses. As it turns out, c_0 is an important parameter in the theory of intergranular friction; it will be expressed later in terms of the bulk properties of the two materials, i.e., the mineral grains and seawater, constituting the sediment. We shall also see that one effect of the friction is to introduce some stiffness into the medium, which raises the actual wave speed above c_0, to a degree that depends on the level of the dissipation.

On introducing a velocity potential, ψ, defined by

$$v = -\frac{\partial \psi}{\partial x}, \tag{4}$$

and combining the above equations, the following wave equation is obtained:

$$\frac{\partial^2 \psi}{\partial x^2} - \frac{1}{c_0^2} \frac{\partial^2 \psi}{\partial t^2} + \frac{(\frac{4}{3}\eta_f + \lambda_f)}{\rho_0 c_0^2} \frac{\partial^3}{\partial t\, \partial x^2} \{h(t) \otimes \psi(t)\} = -S\,\delta(t)\,\delta(x). \tag{5}$$

A driving term representing a planar, impulsive source has been included on the right, in which S is the source strength (with units of length) and $\delta(\cdots)$ is the Dirac delta function. Equation (5) is the cornerstone of the following analysis of wave propagation in a granular medium. The memory function $h(t)$ in Eq. (5) still remains to be specified, to which end the following argument is developed.

It is proposed that, although assumed to possess no skeletal frame, implying zero elastic rigidity, an unconsolidated sediment exhibits behavior resembling that of a visco-elastic solid as a result of the intergranular friction: Loads and deformations are linearly related but the deformation depends on the history of the loading process. In such a material, the response to an instantaneous stress is an instantaneous strain followed by a strain which increases with time. Such behavior in a viscoelastic material is known as ''creep.'' The converse is ''relaxation,'' in which an instantaneous strain gives rise to an instantaneous stress followed by a decrease in stress level with increasing time. Thus creep and relaxation are two sides of the same coin, and in a saturated sediment these phenomena originate in the dissipation occurring at grain-to-grain contacts.

In formulating the material response function, we use as a guide the process of relaxation. A typical relaxation function[35,36] is of the form $(1/t)$ raised to some small power, and accordingly we set

$$h(t) = u(t)\, \frac{t_0^{n-1}}{t^n}, \quad 0 < n < 1, \tag{6}$$

where $u(t)$, the unit step function, ensures that the response of the medium is causal. The scaling constant t_0 in Eq. (6), with units of time, is included explicitly to ensure that $h(t)$ has units of inverse time. [If $h(t)$ were a delta function, representing a viscous medium, it would also have units of $(\text{time})^{-1}$.] Throughout the following analysis t_0 is included explicitly in order to preserve the correct units everywhere. Equation (6) is the proposed memory function that characterizes the dissipation arising from intergranular contacts.

III. SOLUTION OF THE WAVE EQUATION

Equation (5) is a linear, inhomogeneous wave equation that may be solved using standard techniques of Fourier analysis. First, on Fourier transforming both sides of the equation with respect to time, the convolution reduces to a product in the frequency domain, yielding

$$\left[1+j\omega\,\frac{(\frac{4}{3}\eta_f+\lambda_f)}{\rho_0 c_0^2}\,H(j\omega)\right]\frac{\partial^2\Psi(j\omega)}{\partial x^2}+k_0^2\Psi(j\omega) = -S\,\delta(x), \tag{7}$$

where $j=\sqrt{-1}$, ω is angular frequency, wave number $k_0=\omega/c_0$, and upper case letters H and Ψ are transforms of their lower case counterparts.

A second Fourier transform, with respect to distance, x, reduces Eq. (7) to an algebraic equation:

$$\left\{k_0^2-\left[1+j\omega\,\frac{(\frac{4}{3}\eta_f+\lambda_f)}{\rho_0 c_0^2}\,H(j\omega)\right]s^2\right\}\Psi_s=-S, \tag{8}$$

where the wave number s is the transform variable, and the subscript s denotes a transform with respect to x. From Eq. (8), the doubly transformed acoustic field may be written as

$$\Psi_s=\frac{S}{[q^2s^2-k_0^2]}, \tag{9}$$

where

$$q=\left[1+j\omega\,\frac{(\frac{4}{3}\eta_f+\lambda_f)}{\rho_0 c_0^2}\,H(j\omega)\right]^{1/2}. \tag{10}$$

The function $H(j\omega)$ is the temporal Fourier transform of the material response function, $h(t)$, in Eq. (6):

$$\begin{aligned}H(j\omega)&=\int_{-\infty}^{\infty}h(t)\exp-j\omega t\;dt\\&=t_0^{n-1}\int_0^{\infty}t^{-n}\exp-j\omega t\;dt\\&=\frac{\Gamma(1-n)}{(j\omega t_0)^{1-n}},\end{aligned} \tag{11}$$

where $\Gamma(\cdots)$ is the gamma function. The final expression in Eq. (11) is valid for $0<n<1$ but not for n identically equal to zero. It is demonstrated later, in Sec. V on pulse propagation, that if n were (incorrectly) put equal to zero, one consequence would be noncausal behavior in the form of acoustic arrivals before $t=0$. There is no such difficulty provided n, however small, is greater than zero.

With the result in Eq. (11), the radical in Eq. (10) becomes

$$q=[1+\chi_f(j\omega t_0)^n]^{1/2}, \tag{12}$$

where the various constants have been amalgamated into one-dimensionless loss coefficient:

$$\chi_f=\frac{(\frac{4}{3}\eta_f+\lambda_f)\Gamma(1-n)}{\rho_0 c_0^2 t_0}. \tag{13}$$

Equations (9) and (12) underpin the remainder of the discussion. Between them they offer the full solution to the propagation problem, provided the inversions back to the frequency and time domains can be performed.

IV. ATTENUATION AND DISPERSION

From Eq. (9) the frequency-dependent velocity potential is given by the inversion integral over wave number s:

$$\Psi(j\omega)=\frac{S}{2\pi q^2}\int_{-\infty}^{\infty}\left(s^2-\frac{k_0^2}{q^2}\right)^{-1}\exp jsx\;ds. \tag{14}$$

The poles of the integrand, $\pm k_0/q$, lie on either side of the real axis in the complex s plane, in the second and fourth quadrants when $\omega>0$ or the first and third quadrants if $\omega<0$. By integrating around a D in the upper and lower half-planes, respectively, for x positive and negative, the integral can be expressed explicitly:

$$\Psi(j\omega)=-\frac{jS}{2k_0 q}\exp-j\,\frac{k_0|x|}{q}, \tag{15a}$$

or, by substituting for q from Eq. (12) and using $k_0=\omega/c_0$,

$$\Psi(j\omega)=\frac{c_0 S}{2j\omega\sqrt{1+\chi_f(j\omega t_0)^n}}\exp-j\,\frac{\omega|x|}{c_0\sqrt{1+\chi_f(j\omega t_0)^n}}. \tag{15b}$$

According to Eqs. (15), each frequency component in the velocity potential field is an even function of x, as symmetry demands. Since the velocity potential is a real function of time, a further requirement is that $\Psi(j\omega)=\Psi^*(-j\omega)$, where the asterisk denotes a complex conjugate. This condition is clearly satisfied by Eq. (15b). Particular care is taken below to treat negative and positive frequencies correctly.

A standard way of representing the argument of the exponential function in Eq. (15b) is in terms of a complex wave number, k, which may be expressed in the form

$$k=\frac{\omega}{c_p}(1-j\beta), \tag{16}$$

where c_p is the (actual) phase speed of the compressional wave and the magnitude of the dimensionless loss tangent, β, is equal to $1/2Q$. Thus the velocity potential can be written as

$$\Psi=\Psi_0\exp-j\,\frac{\omega}{c_p}|x|\exp-\beta\,\frac{\omega}{c_p}|x|, \tag{17}$$

where Ψ_0 is the amplitude of the wave as given in Eq. (15b). By comparison of Eqs. (15b) and (17), the phase speed, c_p, is given by

$$c_p=c_0\,\mathrm{Re}[1+\chi_f(j\omega t_0)^n]^{1/2}, \tag{18}$$

and the loss tangent is

$$\beta=-\frac{c_p}{c_0}\,\mathrm{Im}[1+\chi_f(j\omega t_0)^n]^{-1/2}. \tag{19}$$

From Eq. (17) the attenuation coefficient, in nepers per unit length, is

$$\alpha_p = \frac{\omega}{c_p}\,\beta, \tag{20}$$

indicating that, if β and c_p were independent of frequency, then the attenuation would scale with the first power of frequency. Notice that in the absence of intergranular friction, when the coefficient χ_f is zero, q in Eq. (12) is real, and the attenuation coefficient vanishes. Another point to note is that β takes the same sign as ω, as shown below, and therefore α_p is always positive or zero, which is a necessary condition if the waveform in Eq. (17) is not to diverge. Equations (18) and (19) are generally valid expressions for the phase speed and attenuation in a medium with a memory of the form shown in Eq. (6).

Considering now the special case when n much less than unity, the term $(j\omega t_0)^n$ in the expressions for c_p and β may be approximated from the following argument. First,

$$[\mathrm{sgn}(\omega)j]^n = \cos(n\pi/2) + j\ \mathrm{sgn}(\omega)\sin(n\pi/2)$$

$$\approx 1 + \frac{jn\pi}{2}\ \mathrm{sgn}(\omega), \tag{21}$$

where, through the signum function sgn(..), explicit account has been taken of the sign of ω; and, second,

$$(|\omega|t_0)^n = \exp[n\ \ln(|\omega|t_0)] = 1 + n\ \ln(|\omega|t_0) + \cdots\ . \tag{22}$$

Hence, on expanding c_p and β [Eqs. (18) and (19)] in Taylor series to first order in the small parameter n, we arrive at the expressions

$$c_p \approx c_0\sqrt{1+\chi_f}\left[1 + \frac{n\chi_f}{2(1+\chi_f)}\ \ln(|\omega|t_0)\right]$$

$$= c_0\sqrt{1+\chi_f}\left[1 + \frac{2\beta_p}{\pi}\ \ln(|\omega|t_0)\right] \tag{23}$$

and

$$\beta \approx \beta_p\ \mathrm{sgn}(\omega), \tag{24a}$$

where the magnitude of β is

$$\beta_p = \frac{n\pi\chi_f}{4(1+\chi_f)}. \tag{24b}$$

The dispersive term in Eq. (23) is proportional to the frequency-independent loss tangent β_p, and is in complete agreement with O'Donnell *et al.*[20] and Horton,[19] who obtained an identical relationship between the c_p and β_p on the basis of the Kronig–Kramers relationships without considering a specific loss mechanism. Although β_p shows considerable variation between sediments (see Fig. 11), it usually takes a value in the range 0.001–0.02. The associated frequency dependence exhibited by the phase speed is very weak, falling between 0.015% and 3% per decade. In most circumstances this is entirely negligible, that is, the phase speed is satisfactorily approximated by the constant value

$$c_p = c_0\sqrt{1+\chi_f}. \tag{25}$$

However, as discussed later in Sec. VI, the dispersive term in the phase speed, although small, cannot be neglected when pulse propagation is being considered.

Equation (25) is an interesting result, which states that the presence of intergranular friction raises the wave speed above the value c_0 determined by the bulk modulus and bulk density of the medium [Eq. (3b)]. In effect, the friction at the intergrain contacts introduces some stiffness into the material, as a result of which the sound speed is increased. By comparing Eq. (25) with the expression for the compressional wave speed in an elastic medium, it is evident that the dissipation coefficient χ_f plays the role of a rigidity modulus. It must be emphasized, however, that the rigidity expressed in Eq. (25) is a purely frictional effect and is not associated with elastic rigidity of the material, which has been taken as zero.

Since the loss tangent β_p in Eq. (24) and the phase speed c_p in Eq. (25) are independent of frequency, it follows that, to this level of approximation, the attenuation coefficient, α_p, in Eq. (20) scales with the first power of frequency:

$$\alpha_p = \frac{|\omega|\beta_p}{c_p} = \frac{|\omega|n\pi\chi_f}{4c_0(1+\chi_f)^{3/2}}. \tag{26}$$

According to this result, intergranular friction gives rise to an attenuation that scales as the first power of frequency, and the associated logarithmic dispersion is expressed by Eq. (23). These features of the field are consistent with the observed geoacoustic properties of many unconsolidated marine sediments. Equation (26) in particular is a result we have been seeking, that is to say, an expression for the attenuation coefficient that matches the characteristic attenuation exhibited by saturated sediments. Two essential features of the theory give rise to this result: a dissipation term in the wave equation that takes account of hysteresis in the medium through a temporal convolution between a material memory function and the particle velocity [Eq. (1)]; and a specific form for the memory function that is characteristic of (dissipative) relaxation in granular media [Eq. (6)].

To establish the precision of Eqs. (24)–(26), consider the term $(\omega t_0)^n$, which is where the dependence on frequency originates in c_p and β [Eqs. (18) and (19)]. For a value of $n = 0.05$, which is representative of many sediments, the variation in $(\omega t_0)^n$ is 12% per decade of frequency. This small frequency dependence translates into an even weaker effect on the phase speed and the attenuation coefficient, as illustrated in Fig. 1, where the exact expressions for c_p and α_p in Eqs. (18)–(20) are plotted over 12 decades of the dimensionless frequency (ωt_0). For comparison, the approximate expressions in Eqs. (25) and (26) are included in the figure. The values of $n = 0.05$ and $\chi_f = 0.25$ used in generating these curves correspond to a medium-loss sediment (i.e., $\beta_p \approx 0.008$ or equivalently a bottom loss of 0.2 dB/m kHz). The exact curves are almost the same as the approximate forms over the huge frequency range in Fig. 1, confirming the validity of Eqs. (24)–(26) in predicting a loss tangent, β_p, that is independent of frequency, an attenuation coefficient, α_p, that scales with the first power of frequency, and a phase speed that shows a very low level of dispersion.

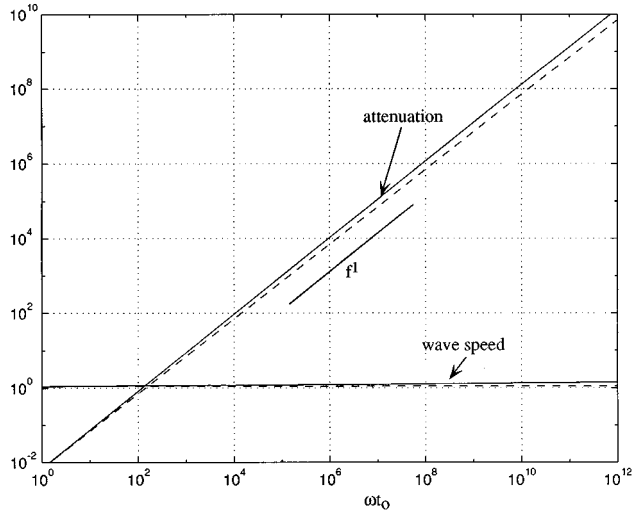


FIG. 1. Normalized attenuation ($\alpha c_0 t_0$) and phase speed (c/c_0) for a material showing intergranular friction. The solid lines are from the exact expressions in Eqs. (18)–(20) and the dashed lines are from the approximations in Eqs. (25) and (26). The curves were computed with $n=0.05$ and $\chi_f=0.25$, corresponding to $\beta_p \approx 0.008$.

V. OBSERVATIONS OF DISPERSION

Wingham[21] has reported a set of careful, laboratory measurements of compressional-wave dispersion and attenuation in gas-free, saturated medium sand. In his Fig. 4 he presents five data sets showing the relative change of phase speed over a frequency range from 100 to 350 kHz. Although he does not state it explicitly, the dispersion he observed in each case is easily calculated by reading from his curves: 1.27%, 0.81%, 1.00%, 0.92%, and 1.08%, the mean of which is 1.02%, per decade of frequency.

The attenuation that Wingham measured in the material scales with the first power of frequency, showing a loss tangent $\beta_p \approx 0.0065$ (equivalent to 0.16 dB/m kHz). When this value of β_p is substituted into our Eq. (23), the theoretically predicted dispersion is found to be 0.95% per decade, in close agreement with the mean value of 1.02% from Wingham's measurements. Dispersion at such a low level is very difficult to measure, especially *in situ*, which could account for the fact that it was not observed by other investigators.[17,18]

VI. PULSE PROPAGATION AND CAUSALITY

It has recently been demonstrated[37] that, in dealing with transient propagation in dissipative media, it is crucial to treat all frequency components accurately. This is especially true of the higher frequencies if the issue of causality is to be addressed, since it is they that govern the behavior of the pulse at short ranges and close to the origin in time.

The direct approach to obtaining the pulse shape in an unconsolidated sediment would be to perform a Fourier frequency inversion of the field expression in Eq. (17). However, with the exact versions of c_p and β in Eqs. (18) and (19), the Fourier integral is difficult to evaluate. On the other hand, if the high-precision approximations, c_p and β_p in Eqs. (24) and (25), are used instead, the integral becomes tractable but the resulting pulse is noncausal. This violation of causality arises directly from neglecting the tiny dispersive component in Eq. (25) in the Fourier inversion. The subtlety of the characteristic attenuation is well illustrated by this behavior.

To develop a satisfactory solution for the pulse shape, we return to the expression for the doubly transformed field in Eq. (9):

$$\Psi_s = -\frac{Sc_0^2}{\omega^2 - s^2\chi_f(j\omega t_0)^n - s^2c_0^2}, \tag{27}$$

where q has been expressed explicitly. The procedure now is to reverse the order of the Fourier inversions by performing the integration over frequency first, to obtain the time-dependent field associated with each wave number. The inversion integral is

$$\psi_s(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Psi_s \exp j\omega t \, d\omega$$

$$= -\frac{Sc_0^2}{2\pi}\int_{-\infty}^{\infty} \frac{\exp j\omega t}{\omega^2 - s^2\chi_f c_0^2(j\omega t_0)^n - s^2c_0^2}\, d\omega. \tag{28}$$

To evaluate this integral, it is necessary to know the positions of the poles in the integrand, which are given by the roots of the denominator. For the case of interest, in which $0<n\ll 1$, these roots may be determined from a simple perturbation analysis.

The equation to be solved is

$$\omega^2 - s^2\chi_f c_0^2(j\omega t_0)^n - s^2c_0^2 = 0. \tag{29}$$

As a zero-order solution, let n be set identically equal to zero, even though the Fourier transform in Eq. (11) is invalid under this condition. Then Eq. (29) has two real roots, $\pm\omega_0$, where

$$\omega_0 = |s|c_0\sqrt{1+\chi_f} = |s|c_p. \tag{30}$$

Since these poles lie on the real axis, it is evident that, with $n=0$, the inversion integral in Eq. (28) is noncausal in that it predicts acoustic arrivals at times t less than zero. Such nonphysical behavior is perhaps to be expected in view of the breakdown of Eq. (11) when $n=0$.

However, when n is finite but small compared with unity Eq. (11) is valid and, as shown below, the difficulty with noncausal arrivals vanishes. Under this condition, the two roots of Eq. (29) are displaced slightly from $\pm\omega_0$ on the real axis to positions $\omega_\pm$ in the complex ω plane. Taking just the leading-order terms in their real and imaginary parts, these complex roots are

$$\omega_\pm = \pm\omega_0 + jn\,\frac{\pi s^2 c_0^2 \chi_f}{4\omega_0}, \tag{31}$$

a result that is based on the approximations

$$(\pm j)^n = 1 \pm \frac{jn\pi}{2} \quad \text{and} \quad (\omega_0 t_0)^n = 1, \tag{32}$$

which are valid for $0<n\ll 1$. It is clear from Eq. (31) that, with n small but finite, both poles lie in the upper half of the complex ω plane, one in the first quadrant (ω_+) and the other symmetrically placed in the second quadrant (ω_-).

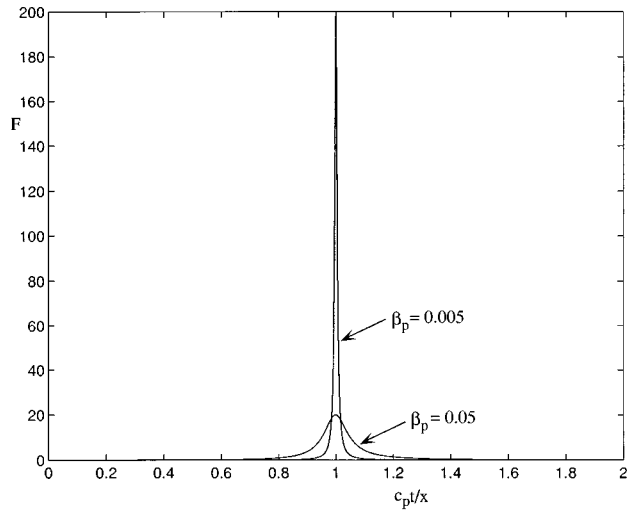


FIG. 2. Normalized velocity pulse, F, in a material showing intergranular friction, for two values of the (frequency-independent) loss parameter, β_p. The curves were computed from Eq. (39).

The inversion integral in Eq. (29) can now be expressed in the form

$$\psi_s(t) = -\frac{Sc_0^2}{2\pi}\int_{-\infty}^{\infty}\frac{\exp j\omega t}{(\omega-\omega_+)(\omega-\omega_-)}\,d\omega, \tag{33}$$

and evaluated directly. For positive times the integration contour must be around the top half-plane, in order to achieve convergence, and for negative times it is around the bottom half-plane. According to Eq. (31) there are no poles in the bottom half-plane, which immediately yields the result

$$\psi_s(t) = -\frac{jSc_0^2}{(\omega_+-\omega_-)}[e^{j\omega_+ t}-e^{j\omega_- t}], \quad t\geqslant 0$$
$$= 0, \quad t<0. \tag{34}$$

Thus every wave number component in the pulse, and hence the pulse itself, is causal. By recognizing that the poles do not lie precisely on the real axis, the difficulty with causality has been removed.

On substituting for the poles $\omega_\pm$ from Eq. (31), the expression in Eq. (34) becomes

$$\psi_s(t) = u(t)Sc_0^2\exp\left(-\frac{n\pi s^2 c_0^2\chi_f}{4\omega_0}\,t\right)\frac{\sin\omega_0 t}{\omega_0}, \tag{35}$$

where $u(t)$ is the unit step function. The particle velocity is obtained directly from Eq. (35) by performing the remaining Fourier inversion integral over wave number:

$$v(t,x) = -\frac{\partial\psi}{\partial x} = -\frac{1}{2\pi}\int_{-\infty}^{\infty} js\psi_s(t)\exp jsx\,ds. \tag{36}$$

With the aid of the expression for ω_0 in Eq. (30), this integral can be written as

$$v(t,x) = u(t)\frac{Sc_0^2}{\pi c_p}$$
$$\times\int_0^{\infty}\exp(-\beta_p sct)\sin(sc_p t)\sin(sx)\,ds, \tag{37}$$

where the loss tangent β_p and the phase speed c_p are given, respectively, by the frequency-independent expressions in Eqs. (24) and (25). Notice that the velocity field is an odd function of x, in accordance with the symmetry of the problem.

On recognizing that the integrand in Eq. (37) can be expressed as the sum of four exponentials, the integration can be performed directly to yield the shape of the velocity pulse:

$$v(t,x) = u(t)\frac{Sc_0^2}{2\pi c_p}\left\{\frac{\beta_p c_p t}{\beta_p^2 c_p^2 t^2+(c_p t-x)^2}\right.$$
$$\left.-\frac{\beta_p c_p t}{\beta_p^2 c_p^2 t^2+(c_p t+x)^2}\right\}$$
$$= u(t)\frac{Sc_0^2}{2\pi c_p x}F\left(\frac{c_p t}{x}\right), \tag{38}$$

where F is the dimensionless function

$$F(y) = \frac{\beta_p y}{(\beta_p y)^2+(y-1)^2}-\frac{\beta_p y}{(\beta_p y)^2+(y+1)^2}. \tag{39}$$

At $t=0$ the velocity waveform and its first derivative with respect to time are both zero. The velocity pulse is a sharply peaked function, as illustrated in Fig. 2 for two values of the loss tangent β_p. With decreasing β_p, the pulse grows taller and narrower, until in the limit of zero dissipation ($\chi_f=0$) it becomes a delta function. Formally, this limiting behavior may be derived with the aid of the identity

$$\lim_{a_1\to 0}\left\{\frac{a_1}{a_1^2+a_2^2}\right\} = \pi\delta(a_2), \tag{40}$$

from which it follows that

$$\lim_{\beta_p\to 0} v(t,x) = \frac{Sc_0}{2}\,\delta(c_0 t-x). \tag{41}$$

This is the correct retarded solution for propagation in a lossless medium. Notice that the velocity pulse in the presence of loss is not a retarded solution, since it varies with x and t individually and not just as $(c_p t-x)$.

The pressure pulse in the granular medium is derived from the velocity pulse by turning to the equation of motion [Eq. (1)], which includes the loss term representing intergranular friction. Although it is possible to derive an expression for the pressure from Eq. (1), the presence of the loss term complicates the analysis significantly. Rather than pursue lengthy algebra, the simple point is made that, since the velocity is causal, the pressure from Eq. (1) is also perfectly well-behaved.

VII. PORE-FLUID VISCOSITY?

It has been suggested by several authors that, in addition to intergranular friction, saturated porous materials such as marine sediments may exhibit viscous losses associated with local and global motions of the pore fluid.[5,27] The following

analysis is presented as an exploratory investigation of the significance of viscosity vis-à-vis intergranular friction in an unconsolidated sediment.

To account for local and global viscous dissipation, the linear equation of motion in Eq. (1), representing saturated sediments exhibiting only intergranular friction, must be modified by the addition of viscous loss terms. In fact, if the effects of local and global losses on wave propagation are assumed to be indistinguishable, a single viscous term is sufficient, allowing the equation of motion for the saturated medium to be written as

$$\rho_0 \frac{\partial v}{\partial t} + \frac{\partial p}{\partial x} - \left(\frac{4}{3}\,\eta_f + \lambda_f\right) \frac{\partial^2 [h(t) \otimes v(t)]}{\partial x^2} - \left(\frac{4}{3}\,\eta + \lambda\right) \frac{\partial^2 v}{\partial x^2} = 0, \tag{42}$$

where the material response function, $h(t)$, is as defined in Eq. (6) and η, λ are the shear and bulk viscosities, respectively, of the pore fluid. It is implicit in Eq. (42) that the viscosity is being regarded as a bulk property of the material, in the same spirit as in the previous treatment of the intergranular friction.

Conservation of mass [Eq. (2)] and the equation of state [Eq. (3)] are unaffected by the dissipation and hence the wave equation is

$$\frac{\partial^2 \psi}{\partial x^2} - \frac{1}{c_0^2} \frac{\partial^2 \psi}{\partial t^2} + \frac{(\frac{4}{3}\,\eta_f + \lambda_f)}{\rho_0 c_0^2} \frac{\partial^3 [h(t) \otimes \psi(t)]}{\partial t\, \partial x^2} + \frac{(\frac{4}{3}\,\eta + \lambda)}{\rho_0 c_0^2} \frac{\partial^3 \psi}{\partial t\, \partial x^2} = -S\,\delta(t)\,\delta(x), \tag{43}$$

where ψ is the velocity potential [Eq. (5)]. It is convenient to amalgamate the two loss terms into one, as follows:

$$\frac{\partial^2 \psi}{\partial x^2} - \frac{1}{c_0^2} \frac{\partial^2 \psi}{\partial t^2} + \frac{(\frac{4}{3}\eta_f + \lambda_f)}{\rho_0 c_0^2} \frac{\partial^3 [\tilde{h}(t) \otimes \psi(t)]}{\partial t \partial x^2} = -S\,\delta(t)\,\delta(x), \tag{44}$$

where the modified material response function, identified by the tilde, is

$$\tilde{h}(t) = b\,\delta(t) + h(t), \quad \text{and} \quad b = \frac{(\frac{4}{3}\eta + \lambda)}{(\frac{4}{3}\eta_f + \lambda_f)}. \tag{45}$$

Clearly, when the relative loss coefficient, b, is zero, corresponding to an absence of viscosity, the linear wave equation in Eq. (44) reduces to the form appropriate to materials exhibiting only intergranular friction [Eq. (5)].

By analogy with Eq. (15a), the solution of Eq. (44) for the frequency dependence of the field may be written immediately as

$$\Psi(j\omega) = -\frac{jS}{2k_0 \tilde{q}} \exp -j\,\frac{k_0 |x|}{\tilde{q}}, \tag{46}$$

where

$$\tilde{q} = \left[1 + j\omega\,\frac{(\frac{4}{3}\eta_f + \lambda_f)}{\rho_0 c_0^2}\,\tilde{H}(j\omega)\right]^{1/2} = [1 + (j\omega t_0)\xi + \chi_f (j\omega t_0)^n]^{1/2}; \quad 0 < n \ll 1. \tag{47}$$

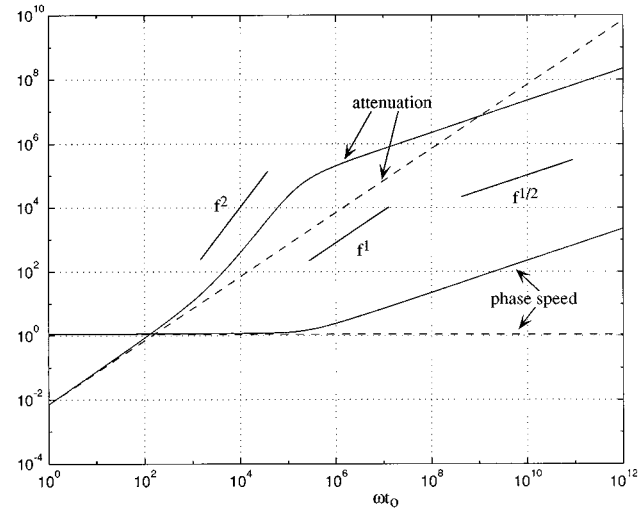


FIG. 3. Normalized attenuation ($\alpha c_0 t_0$) and phase speed (c/c_0). The solid lines, computed from Eqs. (50) to (52) with $\chi_f = 0.25$, $\xi = 10^{-5}$, and $n = 0.05$, corresponding to $\beta_p \approx 0.008$ and $b \approx 4 \times 10^{-5}$, represent a material showing both intergranular friction and viscosity. For reference the dashed lines, from Fig. 1, show the effects of intergranular friction alone.

In these expressions, the temporal Fourier transform of $\tilde{h}(t)$ is

$$\tilde{H}(j\omega) = b + \frac{\Gamma(1-n)}{(j\omega t_0)^{1-n}}, \tag{48}$$

the coefficient ξ representing viscous losses is

$$\xi = \frac{(\frac{4}{3}\eta_f + \lambda_f) b}{\rho_0 c_0^2 t_0} = \frac{b}{\Gamma(1-n)}\,\chi_f, \tag{49}$$

and χ_f, the coefficient for intergranular friction, is as defined in Eq. (13). The phase speed of the wave is

$$\tilde{c}_p = c_0\ \mathrm{Re}(\tilde{q}) = c_0\ \mathrm{Re}[1 + (j\omega t_0)\xi + \chi_f (j\omega t_0)^n]^{1/2} \tag{50}$$

and the attenuation coefficient is

$$\tilde{\alpha}_p = \frac{\omega}{\tilde{c}_p}\,\tilde{\beta}, \tag{51}$$

where

$$\tilde{\beta} = -\frac{\tilde{c}_p}{c_0}\,\mathrm{Im}\left(\frac{1}{\tilde{q}}\right) = -\frac{\tilde{c}_p}{c_0}\,\mathrm{Im}[1 + (j\omega t_0)\xi + \chi_f (j\omega t_0)^n]^{-1/2} \tag{52}$$

is the loss tangent.

Figure 3 shows the phase speed and the attenuation coefficient from Eqs. (50) to (52) as a function of the dimensionless frequency ωt_0. At the lower frequencies, where intergranular friction is the dominant loss mechanism, the attenuation varies linearly with the first power of frequency. However, at intermediate and high frequencies viscosity dominates and the attenuation scales with f^2 and $f^{1/2}$, respectively, as in a viscous fluid. The transition frequency, where the predominant form of loss switches from intergranular friction to viscosity of the pore fluid, depends on the magni-

tude of the relative loss parameter b, defined in Eq. (45). In addition to modifying the attenuation, the presence of viscosity introduces relatively high dispersion, which varies with frequency as $f^{1/2}$ in the higher-frequency, viscous regime.

With the parameters used to generate Fig. 3 ($n=0.05$, $\chi_f=0.25$, and $\xi=10^{-5}$), the effects of viscosity are apparent in the attenuation when the dimensionless frequency ωt_0 is greater than several hundred and in the phase speed when $\omega t_0>10^5$. If ξ were increased, representing higher viscosity, the effects of viscous losses on the attenuation and phase speed would cut in at lower frequencies. A smaller value of ξ, corresponding to lower viscosity, would shift the transition frequency upward. It is interesting that at high frequencies ($\omega t_0>10^9$ in Fig. 3) the attenuation due to the combined effect of intergranular friction and viscosity is actually less than that due to friction alone. At these frequencies the viscosity does not have an additive effect on the attenuation but acts as though intergranular friction were absent, even though the two types of dissipation are represented by additive terms in the wave equation. Even if it existed in real materials, however, this reduced high-frequency attenuation, falling below the f^1 curve for intergranular friction, may not be observable because of the onset of other loss mechanisms, notably scattering.[38]

In fact, in real saturated marine sediments the question of whether pore-fluid viscosity has any detectable effect on the attenuation is not entirely resolved. If it were significant then, according to Eq. (50), it should also introduce an appreciable level of dispersion, that is, a level much higher than that due to intergranular friction alone (Fig. 3). However, the experiments of Wingham,[21] in which he measured the attenuation and dispersion in saturated medium sand up to a frequency of 350 kHz, are entirely consistent with the presence of intergranular friction, but show no evidence of effects due to pore fluid viscosity. As discussed in Sec. V, Wingham found an attenuation that scaled with the first power of frequency, and weak logarithmic dispersion characteristic of intergranular friction, not the relatively strong dispersion, varying as $f^{1/2}$, that would be indicative of viscosity.

The absence of dispersion reported by Hamilton[18] also militates against pore-fluid viscosity as a significant loss mechanism in saturated marine sediments. Moreover, in a discussion of attenuation in sandstone saturated with water, White[13] concludes that losses due to fluid viscosity are entirely negligible. On the other hand, in very early measurements on earth materials, Born[39] found that the attenuation in bars of dry shale, limestone, and sandstone was proportional to the first power of frequency, but when small quantities of water were added the power law changed to f^m, where the exponent m lay between 1 and 2, which may have been due to viscous dissipation. As far as unconsolidated marine sediments are concerned, however, the weight of evidence points to the conclusion that pore-fluid viscosity is negligible and that the dissipation is governed by intergranular friction. Accordingly, viscous losses are excluded from the remainder of the discussion.

VIII. WOOD'S EQUATION FOR c_0

A sediment is a two-phase medium consisting of loose mineral grains and seawater. In the absence of dissipation at the grain-to-grain contacts, the wave speed in such a medium would be given by Wood's equation[40] in terms of weighted means of the bulk moduli and densities of the two constituents of the material. In other words, Wood's equation can be used to express the sound speed c_0 as a function of known mechanical properties of the mineral grains and the seawater.

As given in Eq. (3b), the sound speed in the absence of friction between mineral grains depends on the bulk modulus, κ, and bulk density, ρ_0, of the sediment:

$$c_0=\sqrt{\frac{\kappa}{\rho_0}}. \tag{53}$$

Now, the density ρ_0 is the weighted mean

$$\rho_0=N\rho_w+(1-N)\rho_g, \tag{54}$$

where N is the porosity of the medium and the densities of seawater and mineral grains are, respectively, $\rho_w=1024\ \mathrm{kg/m^3}$ and $\rho_g=2700\ \mathrm{kg/m^3}$. Similarly the bulk modulus of the sediment is given by the mean

$$\frac{1}{\kappa}=N\,\frac{1}{\kappa_w}+(1-N)\,\frac{1}{\kappa_g}, \tag{55}$$

where $\kappa_w=2.25\times10^9$ Pa is the bulk modulus of the pore water. A reasonable value for the bulk modulus of the mineral grains is $\kappa_g=1.47\times10^{10}$ Pa, which is the geometric mean of values reported by Stoll[27] (3.6×10^{10} Pa) and Chotiros[41,42] (6×10^9 Pa). By combining Eqs. (53)–(55), the expression for c_0 becomes

$$c_0=\sqrt{\frac{\kappa_w\kappa_g}{[N\rho_w+(1-N)\rho_g][N\kappa_g+(1-N)\kappa_w]}}, \tag{56}$$

which is Wood's equation for the sound speed in the sediment in the absence of intergranular friction.

Equation (56) is a useful result in that it links an important wave parameter, c_0, with a measurable mechanical property of the sediment, the porosity N. In passing, it is emphasized that Wood's equation is not appropriate for estimating the actual wave speed, c_p, in a sediment, since it takes no account of intergranular friction in the material. In fact, Hamilton[43,44] compared Wood's equation with measurements of wave speed versus porosity and found little correspondence between the data and the theoretical expression. An extension of Wood's theory was developed by Gassmann,[13,45,46] who included a frame bulk modulus in his argument. Although Gassmann's approach may be relevant to rigid, porous media such as rocks, which possess a skeletal elastic frame, it is difficult to see how it could apply to unconsolidated marine sediments.

IX. RANDOM-PACKING, ROUGH-PARTICLE MODEL OF SEDIMENTS

Hamilton[18,47] has measured correlations between wave properties (sound speed and attenuation) and mechanical properties (density, porosity, and grain size) of surficial marine sediments. To establish the corresponding theoretical re-

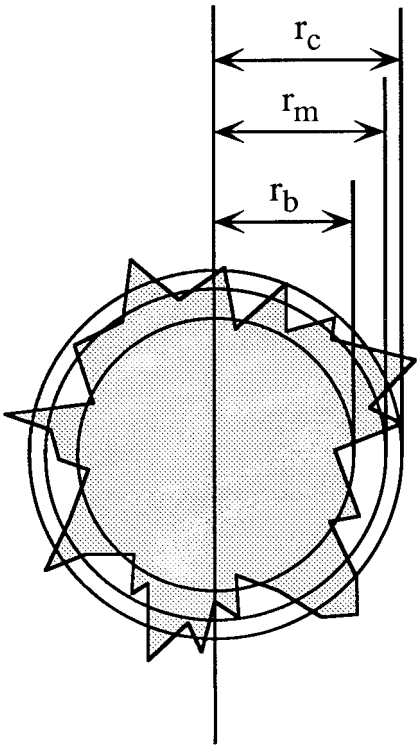


FIG. 4. Schematic of a rough sediment particle. The body radius, r_b, is a measure of the particle size in the absence of the rough surface. The contact radius, r_c, is the effective size of the particle, controlling the packing density in the sediment; and the mean radius, r_m, is intermediate between r_c and r_b. The mean grain diameter of a sediment is identified as $u_g = 2r_b$.

lationships, it is necessary to identify connecting links between the following: (1) density and porosity; (2) intergranular friction and grain size; and (3) porosity and grain size. The first of these pairings is already known and is given by Eq. (54), while the second derives from a simple model of elastic bodies in contact, as discussed below in Sec. XI. The remaining link, between porosity and grain size, which also couples density to grain size through Eq. (54), is now developed on the basis of a random-packing model of the sediment.

Particle roughness plays a central role in the argument, since it is an important factor in determining the porosity and the density of all sediments. To model the porosity, the mineral particles constituting the sediment are treated as rough spheres of uniform size, as illustrated purely schematically in Fig. 4. Measured from the centroid, r_m is the radius of a smooth sphere of volume equal to that of the particle itself. The inner radius, $r_b < r_m$, provides a measure of the body size of the particle, and corresponds approximately to the radius of the sphere that would remain after polishing off the rough surface. The outer radius, $r_c > r_m$, represents the contact radius, that is, on average $2r_c$ is the closest possible approach between the centers of two adjacent sediment grains. According to this description, the roughness prevents the mean surfaces of the particles from making contact, keeping them separated by a finite distance, which, on average, is equal to $2(r_c - r_m)$.

For a sediment in which each grain is in close contact with the surrounding grains, the porosity is the volume fraction of sea water in the medium, which can be expressed as

$$N = 1 - P, \tag{57}$$

where P is the average volume of contiguous grains per unit volume. In the absence of surface roughness, that is, if the grains were smooth spheres, P would be the packing factor, whose value depends on the packing arrangement. For spheres of uniform size, the packing factor is $P = \pi/6 = 0.52$ for simple-cubic symmetry, $P = 0.68$ for a body-centered cubic arrangement, or $P = 0.74$ for both face-centered cubic and hexagonal close-packed structures.[48]

However, for a random packing of smooth spheres, which is more appropriate to sediment particles than any of the regular packing structures, the packing factor has been observed experimentally[49] to be 15% less than the maximum possible value of 0.74 achieved with either of the two close-packed arrangements. Thus a random packing of smooth spheres is described by $P = 0.63$, which is the value adopted for the packing factor of sediment particles in the following discussions. Note that the porosity of a sediment consisting of smooth spherical particles of uniform size would be independent of the sphere radius regardless of the packing arrangement, but in practice such behavior is not observed: The porosity of sediments depends quite strongly on particle size.[18,47]

When the roughness of the spheres is taken into account (Fig. 4), a simple geometrical argument shows that the expression in Eq. (57) for the porosity must be modified as follows:

$$N = 1 - P\,\frac{r_m^3}{r_c^3} = 1 - P\left\{\frac{u_g + 2\Delta}{u_g + 4\Delta}\right\}^3, \tag{58}$$

where $P = 0.63$ is still the packing factor for a random arrangement of smooth spheres of uniform size, and $u_g = 2r_b$ is the body diameter, which is identified with reported measurements of particle size. In formulating Eq. (58), the excursions of the roughness about the mean radius, r_m, have been treated as symmetrical (i.e., the skewness of the distribution is zero). Thus the roughness is characterized by a single statistical parameter, $\Delta = (r_c - r_m) = (r_m - r_b)$, which is identified as the rms roughness height relative to the mean.

If Δ scaled with the body diameter, u_g, then the porosity according to Eq. (58) would still be independent of particle size, which is not a characteristic of real sediments. As shown in Fig. 5, Eq. (58) matches reported porosity versus grain size data[18,47,50] when Δ is taken to be a constant (independent of particle size, u_g) with a value $\Delta = 3\ \mu$m. It can be seen in Fig. 5 that the measurements of sediment porosity decrease with increasing grain size.[18,47,50] The finest-grained materials (clays) show the highest porosity, in the region of $N \approx 0.9$, with the coarse sands exhibiting the lowest value of $N \approx 0.37$. These extreme values of porosity, as observed in actual sediments, correspond to two distinct physical situations. They emerge naturally from Eq. (58), and interestingly enough neither depends on the magnitude of the rms roughness, Δ.

In the limit of large particle size ($u_g \to \infty$), the minimum porosity from Eq. (58) is given by

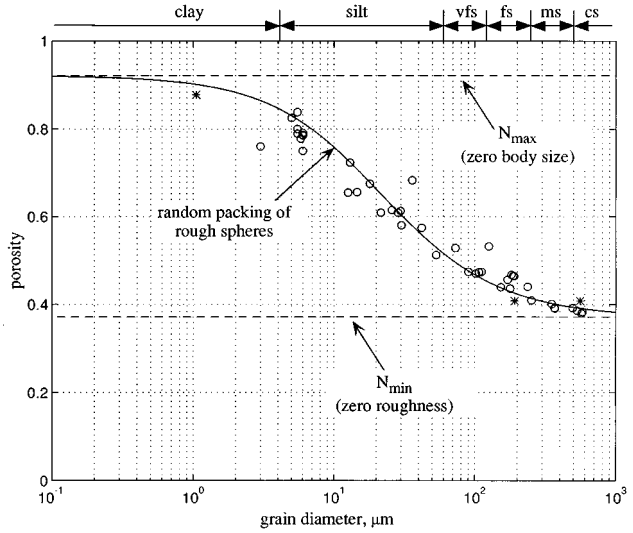


FIG. 5. The porosity as a function of grain diameter. The solid line is from Eq. (58) with $\Delta = 3\ \mu$m and $P = 0.63$; and the dashed lines represent the extreme cases in Eqs. (59) and (60). The symbols are data for saturated marine sediments from: (○), Hamilton (Refs. 18, 47); (*), Richardson and Briggs (Ref. 50). The key to the sediment type is: cs=coarse sand; ms=medium sand; fs=fine sand; vfs=very fine sand.

$$N_{\min} = 1 - P = 0.37, \tag{59}$$

which, physically, corresponds to a random packing of smooth spheres, representative of coarse sand grains whose rms roughness is negligible relative to the body size. At the other extreme of small particle size ($u_g \to 0$), the maximum porosity obtained from Eq. (58) is

$$N_{\max} = 1 - \frac{P}{8} = 0.92, \tag{60}$$

which is characteristic of clay particles that are, in effect, all roughness with zero body size. Between the outer and inner size scales represented by Eqs. (59) and (60), respectively, the shape of the porosity versus grain-size curve does, of course, depend on the magnitude of the roughness, Δ. It can be seen in Fig. 5 that, with the rms roughness set to $\Delta = 3\ \mu$m, as cited above, Eq. (58) provides a very good match to the data in the figure over three decades of grain size. This is an important observation, in view of the fact that the rough, random-packing model is such a simple, one-parameter treatment of sediments. Naturally, it would be desirable to corroborate the size-independent roughness hypothesis, as well as the rms roughness scale, but an independent check is difficult because very few direct measurements of particle roughness have been reported in the literature.

Figures 6 and 7, respectively, show the density versus porosity from Eq. (54) and the density against grain diameter from Eqs. (54) and (58). Included in the figures are measurements of porosity, density, and grain size made by Hamilton,[18,47] and Richardson and Briggs[50] on a large number of surficial sedimentary materials ranging from coarse sand to clay. As with the porosity versus grain size (Fig. 5), the rough-particle, random-packing model of sediments provides a sensible description of the relationship between density and grain size over a range of particle diameters spanning about three decades (clays to coarse sand). Note that the model is based purely on geometry, suggesting that, in surficial sediments, electrostatic and physico-chemical interactions between particles are of little importance in determining the porosity and density of the medium.

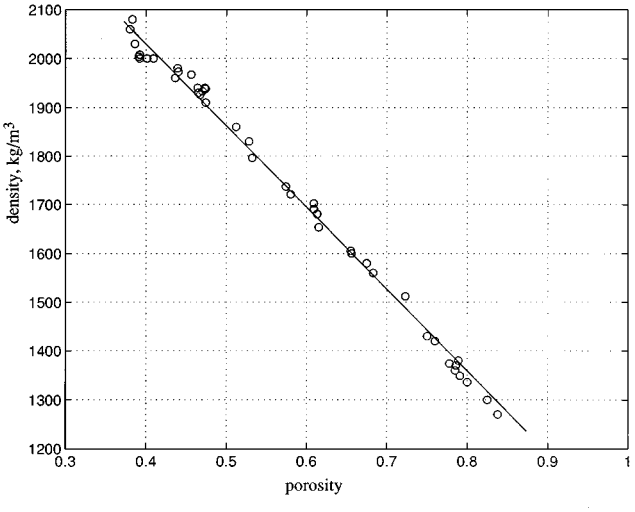


FIG. 6. Plot of density against porosity. The symbols represent data from Hamilton (Refs. 18, 47) and the solid line is the expression in Eq. (54).

If it were not for the fact that it works so well, the unified treatment of sediments embodied in the rough-sphere model may appear to be unduly ambitious. Conventionally, clays, in which the mineral particles are elongated platelets[18] with a very high aspect ratio, are considered to be fundamentally different in structure from the silts and sands, whose grains conform much more closely to the spherical model. Thus although it is reasonable that the rough-sphere model should be representative of the porosity and density of sands and silts, there is a question as to why should it also be applicable to clays?

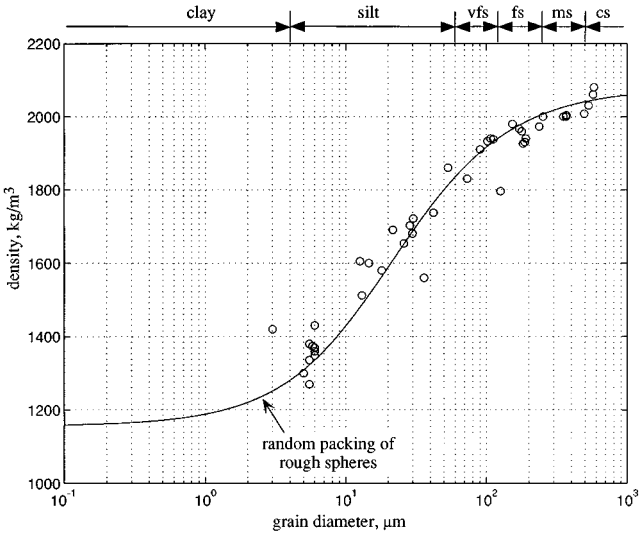


FIG. 7. Sediment density as a function of mean grain diameter. The curved line is from Eqs. (54) and (58), and the symbols represent data from Hamilton (Refs. 18, 47). The key to the sediment type is: cs=coarse sand; ms=medium sand; fs=fine sand; vfs=very fine sand.

In defence of the unified treatment of sediments represented by Eq. (58), the issue of clays is interpreted as follows. It is argued that the geometry of the mineral particles in clay is, in fact, not fundamentally different from that of sand or silt grains. A clay platelet may be thought of as a special, if rather extreme, case of the spherical geometry shown in Fig. 4, in which the body size is much smaller than the rms roughness, Δ. Under this condition, the inherently spherical shape of the particle vanishes, leaving only the roughness, which, if it were to conform to the diagram in Fig. 4, would be statistically isotropic (i.e., the rms roughness would be the same in all radial directions). But a clay platelet is clearly not isotropic. However, if the platelets were randomly orientated in the medium, as they probably are in surficial sediments, given the nature of the depositional process, the effect on the packing structure would be essentially the same as if the individual particles had an isotropic roughness. This would account for the fact that the porosity of surficial clays, like that of the sands and silts, is well represented by the rough-sphere model.

To depths of half a meter or more in the sediment, the porosity of sands and silts is more or less constant, but that of clays tends to decrease by perhaps 10% relative to the surficial value.[51] Such a reduction in porosity of the clays would appear to be at odds with the rough-sphere model [Eq. (58)], since the clay particles half a meter or so beneath the surface are presumably the same size as those at the seawater–sediment interface. The deeper particles, however, may have settled into an arrangement which is no longer isotropic but shows a degree of particle alignment. Even mild polarization would reduce the average center-to-center distance and thus decrease the porosity. Although the rough-sphere model in Fig. 4 could be adapted to account for some anisotropy in the packing structure of the deeper clay particles, this is beyond the scope of the present discussion.

X. POROSITY AND A BI-MODAL GRAIN SIZE DISTRIBUTION

The rough-sphere model as it appears in Eq. (58) is representative of sediments in which the grain-size distribution is unimodal, that is, the particles are fairly uniform in size. As illustrated in Fig. 5, many marine sediments are indeed characterized by a single grain size, but not all. In some sediments, especially the finer silts and the clays, the particle size distribution is bimodal (or perhaps multimodal), the effect of which is to lower the porosity and raise the density of the medium relative to the values that would be observed in the presence of the larger particles alone.[51] For such a sediment, Eq. (58) may be expected to fail.

However, a minor modification to the formulation in Eq. (58) allows for the effect of an admixture of smaller particles on the porosity of the medium. Assuming that the smaller particles simply in-fill the pore spaces formed between the larger grains, that is, they displace their own volume of pore water but have no effect on the random-packing structure of the larger particle species, then the expression for the porosity becomes

$$N = 1 - P\left\{\frac{u_g + 2\Delta}{u_g + 4\Delta}\right\}^3 (1+\gamma), \tag{61}$$

where γ is the volume ratio of the smaller to the larger grain species and u_g is now the mean diameter of the larger grains in the sediment. The remaining symbols in Eq. (61) take the same values as in Eq. (58). To obtain the density of the bimodal sediment, Eq. (61) is substituted into the weighted mean in Eq. (54). Clearly, when $\gamma = 0$, corresponding to a unimodal particle size distribution, Eq. (61) reduces identically to Eq. (58).

Occasionally, a value for γ is reported, along with other geoacoustic data, for sediments with a known bimodal grain-size distribution. For example, γ may be cited indirectly as a percentage of clay in a predominantly silty material. Obviously, γ will never be large enough to make N negative, although in high porosity materials γ may be greater than unity. All the sands exhibit a porosity that is relatively low ($N \leqslant 0.6$), and in these materials the effect of a secondary, in-filling particle species is usually negligible, that is, $\gamma = 0$. This, however, is definitely not the case in some of the silts and the clays, where in-filling of pore space by smaller particles may reduce the porosity and increase the density significantly. For such sediments, it is important to use an appropriate value of γ when estimating the porosity and density of the medium, especially if the model were being used as the basis of a geoacoustic inversion. Having said this, our attention will be focused throughout the remainder of the present discussion on sediments with a unimodal grain-size distribution, for which γ is zero and Eq. (58) holds.

XI. INTERGRANULAR FRICTION AND PARTICLE SIZE

To complete the description of unconsolidated sediments, a relationship must be established which couples the wave properties (sound speed and attenuation) to the mechanical parameters (porosity, density, and grain size). Taking intergranular friction as being the fundamental mechanism underlying wave behavior, and grain size as the independent variable characterizing the mechanical structure of the sediment, the following fractional power-law relationship is proposed:

$$\chi_f = \frac{\mu_c}{\rho_0 c_0^2} = \left(\frac{u_g}{u_0}\right)^{1/3} \frac{\mu_0}{\rho_0 c_0^2}, \tag{62}$$

where u_g is the mean grain diameter (μm) and, as in the phi units of grain size, the reference grain diameter is chosen as $u_0 = 1000$ μm, corresponding to coarse sand. By substituting the first of the expressions in Eq. (62) into Eq. (25), it is clear that the coefficient μ_c is analogous to the rigidity modulus (one of the Lamé parameters) in elasticity theory, but only so far as determining the compressional wave speed is concerned. Thus μ_c is the compressional, frictional-rigidity modulus of the sediment. (As discussed in a later paper, a different frictional-rigidity modulus determines the shear wave speed.) According to Eq. (62), μ_c scales as the cube root of the grain size, the scaling factor being $\mu_0 = 2 \times 10^9$ Pa, which is a material-independent constant whose

value has been established on the basis of a best fit to wave speed versus grain size data (Fig. 8).

It may appear that two unknown parameters have been introduced in Eq. (62), μ_0 and u_0, but it should be clear that they are not independent and that in effect there is only one, namely $\mu_0/(u_0)^{1/3}=2\times10^8$ Pa/$(\mu\text{m})^{1/3}$. The formulation in Eq. (62) avoids the use of a scaling constant with the rather awkward units of pressure/(length)$^{1/3}$.

The expression in Eq. (62) should be valid even for a sediment with a bimodal particle distribution, provided that the smaller particle species merely in-fills the pores formed by the larger particles but plays no part in intergranular dissipation. In such a case, u_g would be the mean diameter of the larger grains.

Equation (62) is suggested by the Hertz theory of deformation of spherical, elastic bodies in contact.[52] When the spheres are pressed together under a force F, a circular area of contact forms with a radius, a, that is proportional to the sphere radius raised to the power of one third:

$$a=\sqrt[3]{\frac{3}{4}\frac{F(1-\sigma^2)}{E}R}, \tag{63}$$

where R is the radius of the spheres, which are taken to be of equal size, and (E,σ) are, respectively, Young's modulus and Poisson's ratio for the elastic material constituting the spheres. The maximum tensile stress acts radially around the circumference of the circle of contact. Taking the two spheres as being adjacent mineral grains in contact, Eq. (62) states that the coefficient of intergranular friction, χ_f, is proportional to the length of the line of highest tensile stress, a distance equal to $2\pi a$, which scales as the cube root of the grain size.

It is implicit in the Hertz theory that the two spherical particles are smooth bodies in contact, whereas in Sec. IX, the mineral grains have been treated as rough particles in the discussion of porosity. These two pictures of the mineral grains may be reconciled by imagining that, in the immediate vicinity of the points of contact, the geometry of the rough particles is locally spherical, suggesting that the cube-root scaling from the Hertz theory in Eq. (62) might be expected to apply. However, the magnitude of the interaction, as measured by the coefficient μ_0 in Eq. (62), is determined, not by the body radius of the particles, but from the rms roughness, that is, the average local radius of curvature at the point of contact, which is independent of grain size. Since a value for μ_0 does not emerge naturally from any available theory of particle-to-particle interaction, it can be determined only from a match to data, which is how the value cited above was obtained.

If Eq. (62) for χ_f as a function of grain size is accepted as a working hypothesis, the sound speed [Eq. (25)] and attenuation [Eq. (24b)] in the sediment take the forms

$$c_p=\sqrt{c_0^2+\left(\frac{u_g}{u_0}\right)^{1/3}\frac{\mu_0}{\rho_0}} \tag{64}$$

and

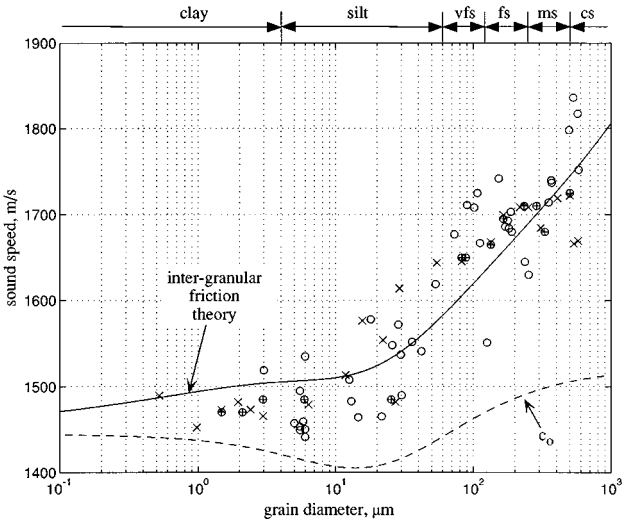


FIG. 8. Wave speed versus mean grain diameter, u_g. The symbols represent data published by Hamilton (Refs. 18, 47) (○), Richardson (Ref. 53) (⊕), and Richardson and Briggs (Ref. 54) (x) and the key to the sediment type is: cs=coarse sand; ms=medium sand; fs=fine sand; vfs=very fine sand. The solid line is c_p from Eq. (64), with c_0 and ρ_0 evaluated from Eqs. (54), (56), and (58). For comparison, the dashed line shows c_0.

$$\beta_p=\frac{n\pi}{4}\frac{u_g^{1/3}}{u_g^{1/3}+w}, \tag{65}$$

where

$$w=\frac{u_0^{1/3}\rho_0 c_0^2}{\mu_0}. \tag{66}$$

These expressions give the wave properties in terms of the grain size, which itself is related to the density, ρ_0, and the porosity, N, through Eqs. (54) and (58). [It should be borne in mind that c_0 is also a function of u_g, as given by Eqs. (56) and (58).] Thus all the parameters of the sediment have been expressed in terms of the grain size, through the simple algebraic expressions developed above.

XII. WAVE SPEED VERSUS MECHANICAL PROPERTIES

Figure 8 shows the wave speed, c_p, plotted as a function of the mean grain size, as evaluated from Eq. (64). The data represented by the symbols in this figure, and also in Figs. 9 and 10, are from tables of measured parameter values published by Hamilton,[18,47] Richardson,[53] and Richardson and Briggs.[54] The value for the scaling coefficient, $\mu_0=2\times10^9$ Pa, cited above in connection with Eq. (62) was obtained by matching the curve to the data in Fig. 8. Otherwise, there are no adjustable parameters available to facilitate the fit. Although there is some spread in the data, it is clear that the theory faithfully follows the trend of the measurements in Fig. 8 over a range of particle sizes spanning at least three decades, representing clays, silts, and all grades of sand. For comparison with c_p, the dashed line in Fig. 8 shows the sound speed, c_0, in the absence of friction, as evaluated from Eqs. (56) and (58).

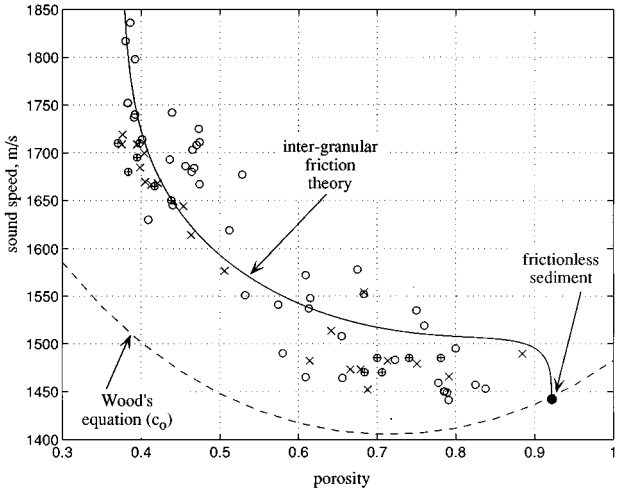


FIG. 9. Wave speed as a function of porosity. The solid line is a plot of Eq. (64) in conjunction with Eqs. (54), (56), and (58), and the symbols represent published data, as identified in the legend to Fig. 8. The dashed line is Wood's equation for c_0 in Eq. (56).

The sound speed versus porosity is shown in Fig. 9, where the solid line represents Eq. (64) in conjunction with Eqs. (54), (56), and (58). Again, the theory of intergranular friction can be seen to provide a very reasonable description of the data. For comparison, Wood's equation for c_0 [Eq. (56)] is included as the broken line in the figure. The only point of intersection of the two theoretical curves corresponds to the ultimate in fine-grained materials, that is to say, a (fictitious) sediment in which the particle (body) size approaches zero and intergranular friction is essentially absent. Elsewhere in the plot the two curves diverge fairly dramatically, but only Eq. (64) for c_p matches the data, thus illustrating the importance of intergranular friction in determining the sound speed in the sediment.

The predicted relationship between sound speed and density is compared with published data in Fig. 10. It can be seen that the solid line, representing Eq. (64) in conjunction with Eqs. (54), (56), and (58), passes through the measured points very nicely. In fact, it is evident from Figs. 8–10 that all the observed correlations between wave speed and the mechanical properties of many sediments are described reasonably by the equations developed in the previous sections. This is interesting for several reasons, not the least being that, with μ_0 and Δ fixed, there are no adjustable parameters in the theory to help match the individual curves to the data.

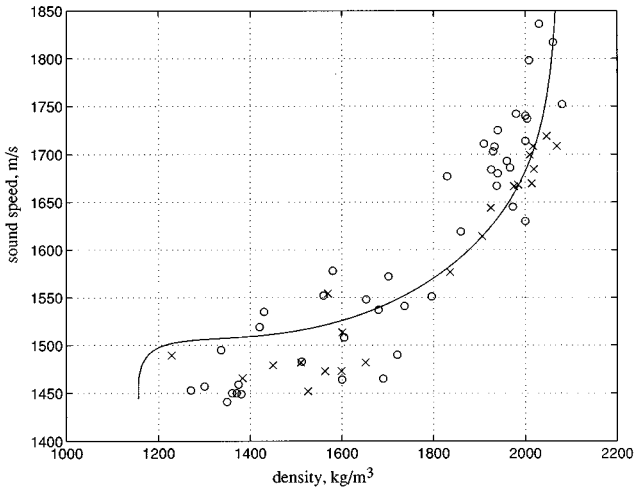


FIG. 10. Sound speed as a function of density. The curved line is a plot of Eq. (64) in conjunction with Eqs. (54), (56), and (58), and the symbols represent data published by Hamilton (Refs. 18, 47) (○), and Richardson and Briggs (Ref. 54) (x).

It is perhaps worth commenting on the spread in the data around the theoretical predictions in Figs. 8–10. The variability in the measurements may be partly due to experimental error, but other factors may also be at work. For instance, in some sediments the bulk modulus, κ_g, and the density, ρ_g, of the mineral particles could be different from the values used in evaluating the theory; and some of the data points in the figures may represent sediments with a bimodal particle size distribution (i.e., $\gamma \neq 0$). Factors such as these could be responsible for the multivalued behavior exhibited by the data in Figs. 8–10. If, for a specific sediment, information is available on the properties of the mineral particles or on the size distribution, it should be used as appropriate to evaluate the theoretical expressions presented above.

XIII. ATTENUATION AND THE MATERIAL MEMORY INDEX, *n*

An important parameter characterizing a sediment is the positive exponent, n, in the memory function [Eq. (6)]. The smaller the value of n, the longer the memory of the material, and *vice versa*. Now, the frequency-independent component of the wave speed, c_p, in Eq. (64) and the mechanical properties of the sediment are independent of n, and thus the correlations shown in Figs. 8–10 are unaffected by the value of n. Only the magnitude of the attenuation, β_p, as given by Eq. (24b), depends on n, a fact which is important in interpreting the highly variable and complicated nature of the level of attenuation observed in nominally similar marine sediments. Some of the mysteries surrounding attenuation in marine sediments are discussed at length by Hamilton.[18]

Incidentally, in a number of his papers, Hamilton expresses the attenuation coefficient in the form

$$\alpha_H = k_H f^m, \tag{67}$$

where α_H is in dB/m, f is frequency in kHz, m is the exponent of frequency, taken here to be unity, and k_H is a constant for a given material. In reporting levels of attenuation in sediments, it has become fairly common practice to cite values of $k_H$in dB/(m kHz), the implicit assumption being that the attenuation scales as the first power of frequency. By comparing Eqs. (26) and (67), it can be seen that k_H depends on the wave speed, c_p, which is usually reported with the attenuation. The conversion relationship between the dimensionless loss tangent, β_p, in Eq. (26) and k_H is

$$\beta_p = 1.83.10^{-5} c_p k_H. \tag{68}$$

The variability of the attenuation in sediments is illustrated in Fig. 11, which shows measurements of the loss tangent, β_p, as a function of the mean grain size, as reported by Hamilton,[18] Richardson,[53] and Richardson and Briggs.[54] The attenuation data are much more scattered than the wave

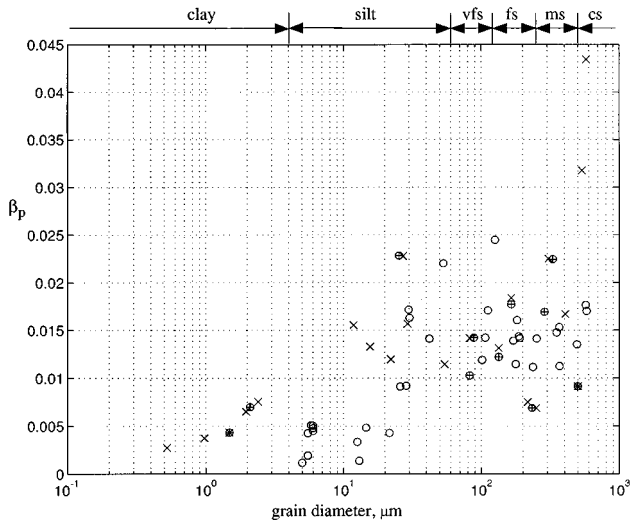


FIG. 11. Plot of loss tangent, β_p, against grain size, u_g. The symbols represent data published by Hamilton (Ref. 18) (○), Richardson (Ref. 53) (⊕), and Richardson and Briggs (Ref. 54) (x). The key to the sediment type is: cs=coarse sand; ms=medium sand; fs=fine sand; vfs=very fine sand.

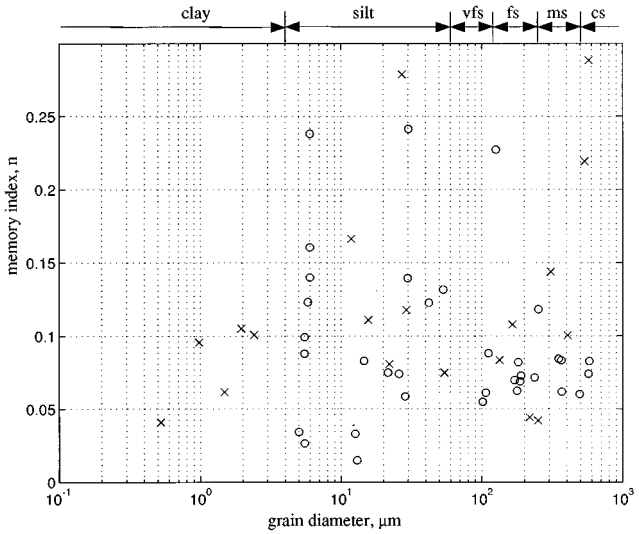


FIG. 13. The memory index n as a function of the grain diameter. The symbols, representing experimental values derived from published data, are as identified in the legend to Fig. 12, and the key to the sediment type is: cs=coarse sand; ms=medium sand; fs=fine sand; vfs=very fine sand.

speed measurements in Figs. 8–10. According to Eq. (24b), the large variability in β_p could originate either with the loss parameter χ_f or with the memory index n. Now, the coefficient χ_f is also important in determining the wave speed [Eq. (25)]; but, as Figs. 8–10 illustrate, the wave speed shows relatively little scatter, following reasonably well-defined trends. This stable behavior suggests that χ_f is itself fairly well-behaved and is not responsible for the variability in β_p. Further support for this conclusion is evident in Fig. 12, which shows χ_f as computed from published measurements of sound speed, porosity, density, and grain size,[18,54] using an inverted form of Eq. (25):

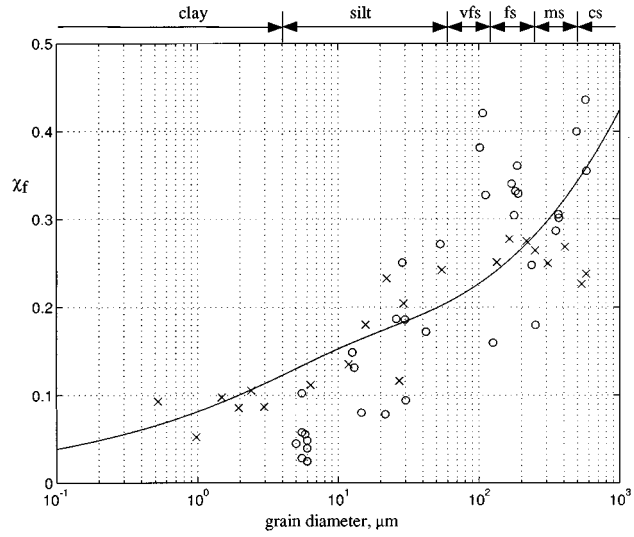


FIG. 12. Plot of the attenuation coefficient, χ_f, versus grain size. The solid line is the theoretical curve from Eq. (62) and the symbols represent experimental values derived from data published by Hamilton (Ref. 18) (○), and Richardson and Briggs (Ref. 54) (x). The key to the sediment type is: cs=coarse sand; ms=medium sand; fs=fine sand; vfs=very fine sand.

$$\chi_f = \frac{c_p^2 - c_0^2}{c_0^2}. \tag{69}$$

To generate the points in the figure from this expression, c_0 was computed from Wood's equation using measured values of density and porosity. The solid line in the figure is the theoretical expression for χ_f in Eq. (62). It can be seen that the values of χ_f derived from data show some scatter about the theoretical curve but not sufficient to account for the observed variability in the attenuation.

If χ_f is not responsible for the scatter in the attenuation data, the culprit must be the memory index n (assuming that the variability is genuine and not an artifact arising from, say, scattering or experimental error). This is confirmed by estimating n from attenuation measurements,[18,54] taken in conjunction with the values of χ_f shown in Fig. 12, using an inverted form of Eq. (24b):

$$n = \frac{4\beta_p}{\pi} \frac{(1+\chi_f)}{\chi_f}. \tag{70}$$

Figure 13 shows the resultant values of n as a function of the grain size. It is evident that n varies wildly, even for sediments with similar grain size. For example, the values of n for the 8 nominally identical fine-slit sediments, with mean grain size around 5–6 μm, range from 0.03 to 0.24, that is, a factor of 8 between the highest and lowest values of n in these otherwise similar materials. (For these particular materials, the reported values of the porosity and bulk density are $N \approx 0.8$ and $\rho_0 \approx 1330$ kg/m^3, respectively.) The mean value of all the points in the Fig. 13 is $\overline{n} = 0.1$, and the standard deviation about the mean is $\sigma_n = 0.06$, which is comparable with the mean, indicative of the high level of scatter.

The widely spread distribution of data points in Fig. 13, following no obvious trend, identifies the principal cause of scatter in the attenuation measurements (Fig. 11) as the exponent, n, of the memory function in Eq. (6). Why should n

show such variability, particularly for macroscopically similar sediments? A plausible reason could be that the length of the memory is very sensitive to the nature of the bonding between mineral grains. This could be affected by the compactness of the sediment, overburden pressure, sharpness of the grains, and perhaps other factors. The actual interparticle forces involved depend on the microstructure and physico-chemical nature of the sediment grains, as discussed by Hamilton,[18] but such issues are beyond the scope of the present discussion.

The question of whether the magnitude of the attenuation in marine sediments, β_p, can be predicted, given macroscopic, mechanical information such as particle size or porosity, is important in connection with applications in geophysics and underwater acoustics. On the basis of the spread of the data in Figs. 11 and 13, it would appear that any such prediction could be considerably in error, since there is little correlation between attenuation and, say, particle size. The memory index n, which determines the loss tangent, β_p, is probably governed by microscopic interactions between mineral grains, rather than the macroscopic, mechanical properties of the sediment. These microscopic forces need to be characterized if satisfactory predictions of the level of attenuation are to be achieved.

XIV. CONCLUDING REMARKS

The acoustic properties of marine sediments are of practical interest to underwater acousticians concerned with the prediction of acoustic propagation loss, and to seismologists, geophysicists and geologists in connection with offshore oil and gas exploration. Acoustic attenuation is central to these investigations and accordingly has been the subject of very extensive research. There is now considerable evidence to support the view that in many unconsolidated marine sediments the attenuation scales as the first power of frequency over some six decades of frequency. A satisfactory explanation of this phenomenon has not been readily forthcoming.

In the unified theory of sediment acoustics developed in this article, the medium is treated as a fluid that supports a previously unexplored form of intergranular dissipation. The characteristic feature of the new loss mechanism is hysteresis, that is, the frictional stress at intergrain contacts exhibits a memory. The linear scaling of attenuation with frequency emerges quite naturally from the analysis, and the compressional wave speed is shown to exhibit logarithmic dispersion, although at such a low level as to be negligible in most but not all circumstances. According to the theory, pulse propagation in the medium is strictly causal, as of course it must be, although a simple (but fallacious) Fourier inversion argument, which neglects the tiny level of dispersion, would seem to suggest otherwise. This apparent violation of causality has fueled many a debate in the literature. Without doubt the inter-relationships between attenuation, dispersion, and causality in the context of sediments are subtle but not pathological.

To augment the acoustic theory, a new model of the mechanical properties of sediments is developed, based on the idea of randomly packed, rough mineral grains. When the mechanical model is coupled to the wave theory through a simple relationship between a frictional coefficient and the grain size, expressions are obtained that relate the acoustic properties (wave speed and attenuation) to the mechanical properties (grain size, density, and porosity). As far as the wave speed is concerned, these theoretical correlations show very nice agreement with published data.

The corresponding relationships for the magnitude of the attenuation cannot be easily compared with data, because measurements of attenuation show a high degree of variability for nominally identical sediments. In other words, the level of attenuation does not correlate well with the mechanical properties of the sediment. The new theory has been used to trace the cause of the scatter in the attenuation data back to the material memory function. The level of attenuation appears to be governed by the length of the memory, which could be very sensitive to interparticle bonding forces. Such forces, acting on a microscopic scale, may be very different in otherwise similar materials, due to factors such as compactness, overburden pressure and physico-chemical effects. If this interpretation is correct, it will be difficult to make a prediction of the level of attenuation in a sediment with any degree of confidence until the connection between particle bonding and the material memory function has been established.

In this article, the analysis has concentrated on compressional wave propagation in an unbounded sediment. Shear waves have been neglected in the analysis, even though saturated marine sediments are known to support shear, albeit with wave speeds much slower than those in consolidated rocks. By extending the theoretical arguments developed here in connection with compressional waves, the properties of shear waves in unconsolidated sediments will be established in a second paper on the geoacoustic properties of granular materials. A third paper in the series will address interface wave propagation at the sediment–seawater boundary, on the assumption that the rigidity in the sediment arises solely from inter-granular dissipation.

ACKNOWLEDGMENTS

I have benefited from many discussions with Dr. Grant Deane on attenuation in marine sediments. Dr. Michael Richardson very kindly reviewed the paper in great detail and made many suggestions that lead to a much improved treatment of the mechanical properties of sediments. This research was supported by the Office of Naval Research under Contract No. N00014-91-J-1118, for which I am grateful.

[1] A. C. Kibblewhite, ''Attenuation of sound in marine sediments: A review with emphasis on new low frequency data,'' J. Acoust. Soc. Am. **86**, 716–738 (1989).
[2] R. D. Mindlin, ''Compliance of elastic bodies in contact,'' J. Appl. Mech. **16**, 259–268 (1949).
[3] M. N. Toksöz, D. H. Johnston, and A. Timur, ''Attenuation of seismic waves in dry and saturated rocks, I. Laboratory measurements,'' Geophysics **44**, 681–690 (1979).
[4] D. H. Johnston, M. N. Toksöz, and A. Timur, ''Attenuation of seismic waves in dry and saturated rocks, II. Mechanisms,'' Geophysics **44**, 691–711 (1979).
[5] K. W. Winkler and A. Nur, ''Seismic attenuation: Effects of pore fluids and frictional sliding,'' Geophysics **47**, 1–15 (1982).

[6] E. L. Hamilton, ''Sound attenuation as a function of depth in the sea floor,'' J. Acoust. Soc. Am. **59**, 528–535 (1976).
[7] E. L. Hamilton, ''Acoustic properties of sediments,'' in *Acoustics and the Ocean Bottom*, edited by A. Lara-Saenz, C. Ranz Cuierra, and C. Carbo-Fité (Consejo Superior de Investigacions Investigacions Cientificas, Madrid, 1987), pp. 3–58.
[8] L. Bjørnø, ''Features of the linear and non-linear acoustics of water-saturated marine sediments,'' Technical University of Denmark, Report number AFM 76-06 (1976).
[9] M. N. Toksöz and D. H. Johnston, ''Seismic Wave Attenuation,'' in *Geophysics Reprint Series, No. 2* (Society of Exploration Geophysicists, Tulsa, OK, 1981).
[10] E. Strick, ''Predicted pedestal effect for pulse propagation in constant-*Q* solids,'' Geophysics **35**, 387–403 (1970).
[11] R. W. Morse, ''Acoustic propagation in granular media,'' J. Acoust. Soc. Am. **24**, 696–700 (1952).
[12] J. W. S. Baron Rayleigh, *The Theory of Sound* (Dover, New York, 1945).
[13] J. E. White, *Underground Sound: Application of Seismic Waves* (Elsevier, Amsterdam, 1983).
[14] G. K. Batchelor, *An Introduction to Fluid Dynamics* (Cambridge U.P., Cambridge, 1967).
[15] P. M. Morse and K. U. Ingaard, *Theoretical Acoustics, International Series in Pure and Applied Physics* (McGraw-Hill, New York, 1968).
[16] C. W. Horton, Sr., ''Dispersion relationships in sediments and sea water,'' J. Acoust. Soc. Am. **55**, 547–549 (1974).
[17] F. J. McDonal, F. A. Angona, R. L. Mills, R. L. Sengbush, R. G. Van Nostrand, and J. E. White, ''Attenuation of shear and compressional waves in Pierre shale,'' Geophysics **23**, 421–439 (1958).
[18] E. L. Hamilton, ''Compressional wave attenuation in marine sediments,'' Geophysics **37**, 620–646 (1972).
[19] C. W. Horton, Sr., ''Comment on 'Kramers–Kronig relationship between ultrasonic attenuation and phase velocity' [J. Acoust. Soc. Am. **69**, 696–701 (1981)],'' J. Acoust. Soc. Am. **70**, 1182 (1981).
[20] M. O'Donnell, E. T. Jaynes, and J. G. Miller, ''Kramers–Kronig relationship between ultrasonic attenuation and phase velocity,'' J. Acoust. Soc. Am. **69**, 696–701 (1981).
[21] D. J. Wingham, ''The dispersion of sound in sediment,'' J. Acoust. Soc. Am. **78**, 1757–1760 (1985).
[22] E. L. Hamilton, ''Geoacoustic models of the sea floor,'' in *Physics of Sound in Marine Sediments* (Plenum, New York, 1974), pp. 181–221.
[23] E. L. Hamilton, ''Geoacoustic modeling of the sea floor,'' J. Acoust. Soc. Am. **68**, 1313–1336 (1980).
[24] 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).
[25] 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).
[26] M. A. Biot, ''Mechanics of deformation and acoustic propagation in porous media,'' J. Appl. Phys. **33**, 1482–1498 (1962).
[27] R. D. Stoll, *Sediment Acoustics, Lecture Notes on Earth Sciences* (Springer-Verlag, Berlin, 1989).
[28] J. B. Walsh, ''Seismic wave attenuation in rock due to friction,'' J. Geophys. Res. **71**, 2591–2599 (1966).
[29] W. P. Mason, ''Internal friction mechanism that produces an attenuation in the Earth's crust proportional to frequency,'' J. Geophys. Res. **74**, 4963–4966 (1969).
[30] M. D. Verweij, ''Modeling space-time domain acoustic wave fields in media with attenuation: The symbolic manipulation approach,'' J. Acoust. Soc. Am. **97**, 831–843 (1995).
[31] M. D. Verweij, ''Transient acoustic wave modeling: Higher-order Wetnzel–Kramers–Brillouin–Jeffreys asymptotics and symbolic manipulation,'' J. Acoust. Soc. Am. **92**, 2223–2238 (1992).
[32] A. T. de Hoop, ''Acoustic radiation from an impulsive point source in a continuously layered fluid—An analysis based on the Cagniard method,'' J. Acoust. Soc. Am. **88**, 2376–2388 (1990).
[33] A. T. de Hoop, ''Similarity analysis for the elastic wave motion in a dissipative solid under a global relaxation law,'' in *IUTAM Symposium—Nonlinear Waves in Solids*, edited by J. L. Wegener and F. R. Norwood (American Society of Mechanical Engineers, Fairfield, NJ, 1995), p. 77–82.
[34] G. B. Deane, ''Internal friction and boundary conditions in lossy fluid sea beds,'' J. Acoust. Soc. Am. **101**, 233–240 (1997).
[35] J. D. Achenbach, *Wave Propagation in Elastic Solids, Applied Mathematics and Mechanics* (North-Holland, Amsterdam, 1975).
[36] M. A. Biot, ''Generalized theory of acoustic propagation in porous dissipative media,'' J. Acoust. Soc. Am. **34**, 1254–1264 (1962).
[37] M. J. Buckingham, ''Acoustic pulse propagation in dispersive media,'' in *New Perspectives on Problems in Classical and Quantum Physics* (Gordon and Breach, New York, in press).
[38] K. W. Winkler, ''Frequency dependent ultrasonic properties of high-porosity sandstones,'' J. Geophys. Res. **88**, 9493–9499 (1983).
[39] W. T. Born, ''The attenuation constant of earth materials,'' Geophysics **6**, 132–148 (1941).
[40] A. B. Wood, *A Textbook of Sound* (Bell and Sons, London, 1964).
[41] J. C. Molis and N. P. Chotiros, ''A measurement of grain bulk modulus of sands,'' J. Acoust. Soc. Am. **91**, 2463 (A;1992).
[42] N. P. Chotiros, ''Biot model of sound propagation in water-saturated sand,'' J. Acoust. Soc. Am. **97**, 199–214 (1995).
[43] E. L. Hamilton, ''Prediction of deep-sea sediment properties: state-of-the-art,'' in *Deep-Sea Sediments: Physical and Mechanical Properties*, Vol. 2, edited by A. L. Inderbitzen (Plenum, New York, 1974), pp. 1–43.
[44] E. L. Hamilton, ''Elastic properties of marine sediments,'' J. Geophys. Res. **76**, 579–604 (1971).
[45] F. Gassmann, ''Über die elastizität poröser medien,'' Vierteljahrsscher. Naturforsch. Ges. Zürich, **96**, 1–23 (1951).
[46] F. Gassmann, ''Elastic waves through a packing of spheres,'' Geophysics **16**, 673–685 (), and **18**, 269 (1951).
[47] E. L. Hamilton, ''Sound velocity and related properties of marine sediments, North Pacific,'' J. Geophys. Res. **75**, 4423–4446 (1970).
[48] P. C. Gasson, *Geometry of Spatial Forms* (Horwood, Chichester, 1983).
[49] O. K. Rice, ''On the statistical mechanics of liquids, and the gas of hard elastic spheres,'' J. Chem. Phys. **12**, 1–18 (1944).
[50] M. D. Richardson and K. B. Briggs, ''*In situ* and laboratory geoacoustic measurements in soft mud and hard-packed sand sediments: Implications for high-frequency acoustic propagation and scattering,'' Geo-Marine Lett. **16**, 196–203 (1996).
[51] M. D. Richardson, personal communication (1997).
[52] S. Timoshenko and J. N. Goodier, *Theory of Elasticity* (McGraw-Hill, New York, 1951).
[53] M. D. Richardson, ''Spatial variability of surficial shallow water sediment,'' in *Ocean-Seismo Acoustics*, edited by T. Akal and J. M. Berkson (Plenum, New York, 1986), pp. 527–536.
[54] M. D. Richardson and K. B. Briggs, ''On the use of acoustic impedance values to determine sediment properties,'' in Proc. Inst. Acoust. **15**, Pt. 2, 15–24 (1993).

Wide-band source localization in the presence of internal waves

Paul J. Book and L. W. Nolte
Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina 27708

(Received 1 November 1996; revised 21 July 1997; accepted 24 July 1997)

The effects of internal waves on low-frequency (25-Hz) wide-band source localization performance are investigated at a very long range (1000 km). Sound-speed perturbations induced by internal waves obeying the Garrett–Munk spectral model are incorporated into the normal mode solution to the wave equation using the adiabatic approximation. Source range estimates are limited by very random internal waves and do not improve with increasing SNR. In an attempt to combat this limitation due to environmental uncertainty, processors can benefit from the spectrum of the acoustic source. According to the hybrid Cramer–Rao lower bound (CRLB), significant improvements on range estimates are possible when using a coherent wide-band acoustic source with bandwidth of 10 Hz on the 25-Hz source. The CRLB provides valuable analytical insight into the limits on source range estimates, but the optimal signal processing algorithm is necessary to compute the actual limits on source localization. Using a coherent wide-band source with many discrete frequencies spanning the bandwidth, significant gains in range estimates are achieved by optimally incorporating this signal spectrum into the localization algorithm *a priori*. © *1997 Acoustical Society of America.* [S0001-4966(97)02811-7]

PACS numbers: 43.30.Wi, 43.60.Gk [SAC-B]

INTRODUCTION

This paper quantifies the performance and limitations of low-frequency (25-Hz) wide-band source localization in the presence of internal-wave variability at a very long range (1000 km) in the deep ocean. Environmental mismatch is known to have a significant effect on the performance of source localization using matched-field processing. Internal waves cause most of the ocean variability on time scales under 12 h. Colosi and Flatte have explored the effect internal waves cause on modal coupling for multi-megameter range wide-band acoustic propagation.[1] By incorporating environmental uncertainty into the matched-field processing technique *a priori*, current processors are made more robust and perform better in mismatched environments.[2–4] Internal-wave characteristics have been optimally incorporated *a priori* into a Bayesian source localization processor, known as the optimum uncertain field processor (OUFP).[5] Even with the internal-wave variability optimally incorporated *a priori*, all processors are limited in very random internal-wave environments. In an attempt to combat this limitation due to environmental uncertainty, processors can benefit from the spectrum of the acoustic source.

In this paper, the limits of wide-band source localization are computed by optimally incorporating environmental uncertainty and source spectrum *a priori* into the OUFP. Source range estimates are limited by very random internal waves and do not improve with increasing SNR; consequently, this work focuses on determining the necessary source spectrum to improve source range estimates limited by internal-wave variability. The optimum signal processing algorithm computes these limits for the entire range of signal-to-noise ratios (SNR). The localization errors are local in high SNR; consequently, the Cramer–Rao lower bound (CRLB) can be used to predict the theoretical limits on source range estimates to provide insight on limits of range estimates for various source spectra.

A general description of the localization problem and the specific models and processors that are implemented are discussed in Sec. I. Once the models and processors are implemented, analytical and numerical simulations are performed in Sec. II. Finally, Sec. III concludes with a discussion of the results.

I. PROBLEM DESCRIPTION

Very long range wide-band source localization is performed in a deep ocean containing an uncertain internal wave (Fig. 1). The ocean model is described in Sec. I A and the sound-speed perturbations caused by an internal wave obeying the Garrett–Munk spectrum are discussed in Sec. I. The *a priori* source spectrum necessary to improve wide-band source localization is determined. The acoustic wave propagation of this source is simulated by the normal mode solution to the wave equation using KRAKEN and the adiabatic approximation.[6,7] The details of incorporating internal-wave induced sound speed perturbations into the normal mode solution to the wave equation using the adiabatic approximation are in Sec. I B. A fully spanning vertical receiver array is used to sample the acoustic pressure field. The optimum uncertain field processor (OUFP) computes the localization limits of this acoustic source in this ocean containing an uncertain internal wave. The Bartlett processor uses only the mean ocean and is implemented to determine the consequences of ignoring the internal wave. The matched-field processor is also implemented and provides the upper bound on source localization when the ocean environment is known. Section I C contains details of the processors. The Cramer–Rao lower bound (CRLB) does not provide an algorithm to localize the acoustic source, but rather analyti-

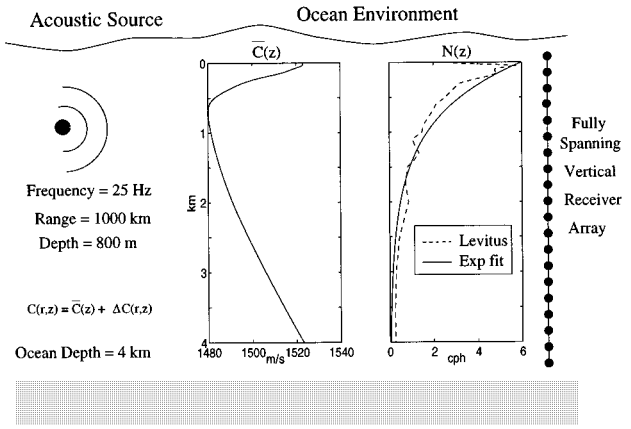


FIG. 1. Deep ocean localization problem with average sound velocity profile, $\overline{C}(z)$, and buoyancy frequency, $N(z)$.

cally computes potential performance bounds on source range estimates. The analytical computations for the CRLB are discussed in Sec. I D.

A. Ocean model

The ocean is range dependent with a constant depth of 4 km. The range-dependent sound velocity profile model consists of a known range-independent average plus a range-dependent perturbation due to an internal wave $[C(r,z) = \overline{C}(z) + \Delta C(r,z)]$. The average sound velocity profile, $\overline{C}(z)$, is computed by converting temperature and salinity profiles from Kauai (159.46 °W, 22.44 °N) to Point Sur (122.00 °W, 36.25 °N) using Del Grosso's relationship (Fig. 1). The temperature and salinity profiles are provided by Levitus and NOAA.[8] The average buoyancy frequency, $N(z) = \sqrt{-(g/\rho)(\partial\rho/\partial z)}$, is also computed using the oceanographic toolbox from Woods Hole Oceanographic Institution. An exponential fit $[N(z) = N_0 e^{-z/B}$, where $N_0 = 6$ cycles/h is the extrapolated surface frequency and $B = 800$ m] is applied to the average buoyancy frequency.

1. Garrett–Munk internal-wave model

Perturbations in density stratification cause internal waves. Garrett and Munk have developed an internal-wave spectrum which appears to be a first-order description of deep water internal waves.[9,10] Reviewing this model briefly, the vertical displacements of the propagating modes of internal waves are described by

$$\frac{\partial^2 W(j,k,z)}{\partial z^2} + k^2\left(\frac{N^2(z)-\omega^2(j,k)}{\omega^2(j,k)-\omega_i^2}\right) W(j,k,z) = 0, \qquad (1)$$

where j is the discrete mode number of the internal wave ($j = 1,2,3,\ldots,j_n$), k [rad/m] is the horizontal spatial frequency of the internal wave, and z [m] is depth. $W(j,k,z)$ [m] is the magnitude of the vertical displacement with boundary conditions of $W(j,k,0) = W(j,k,z_{\text{bottom}}) = 0$. Also, ω_i [cycles/h] $= 2\sin(\text{latitude})/24$ cycles/h is the inertial (temporal) frequency and is included to account for the Coriolis force. When the latitude is 30°, $\omega_i = 1/24$ cycles/h. $\omega(j,k)$ [cycles/h] is the (temporal) frequency of the internal-wave mode. These internal-wave eigenmodes are orthogonal and normalized by

$$\int_0^{z_{\text{bottom}}} (N^2(z)-\omega_i^2) W(j,k,z) W(j',k,z)\,dz = \delta(j-j'). \qquad (2)$$

After projecting the total internal-wave displacement onto a two-dimensional ocean slice, the normalized vertical displacement, $\zeta(r,z,t)$, is expressed as a superposition of modes with random complex amplitudes:

$$\zeta(r,z,t) = \zeta_0 N_\sigma \sum_{j=1}^{j_n} \sum_{k=k_1}^{k_n} \sum_{\theta=\theta_1}^{\theta_n} W(j,k,z)\sigma(j,k) \times \text{Re}\{AIW(j,k,\theta) e^{i[k\cos(\theta)r - \omega(j,k)t]}\}, \qquad (3)$$

where the complex amplitudes of the internal-wave modes, $AIW(j,k,\theta)$, are independent complex Gaussians with zero mean and unit variance [i.e., $\overline{AIW}(j,k,\theta) = 0$ and $\langle AIW(j,k,\theta)^2\rangle = 1$]. The angles between the internal-wave modes and the ocean slice, θ, have independent uniform distributions $[0 \leqslant \theta < 2\pi]$. The internal-wave field is to have a specific rms displacement, $\zeta_0 = \zeta(z_{\text{ref}})_{\text{rms}} = 7.3$ m, at a depth z_{ref} where $N(z_{\text{ref}}) = 3$ cycles/h. Using the average buoyancy frequency along the path from Kauai to Point Sur, $z_{\text{ref}} = 550$ m and $N_0 B = 6$ cycles/h$\times$800 m. To properly normalize the vertical displacement

$$N_\sigma^{-2} = \sum_{j=1}^{j_n} \sum_{k=k_1}^{k_n} \sum_{\theta=\theta_1}^{\theta_n} W^2(j,k,z_{\text{ref}})\sigma^2(j,k)\,\frac{1}{2}, \qquad (4)$$

$$\sigma^2(j,k) = \int_{k-\Delta}^{k+\Delta} \frac{j_*}{j^2 + j_*^2} \frac{\hat{k}^2 k(j)}{(\hat{k}^2 + k^2(j))^2}\,d\hat{k},$$

$$k(j) = \frac{\pi j \omega_i}{N_0 B}, \qquad (5)$$

and $j_* = 3$. The Garrett–Munk spectrum is normalized by

$$\sum_{j=1}^{j_n} \sum_{k=k_1}^{k_n} \sum_{\theta=\theta_1}^{\theta_n} \sigma^2(j,k) = 1. \qquad (6)$$

The internal-wave energy is distributed according to the Garrett–Munk spectrum. These internal-wave displacements are converted into a perturbation in sound speed by[10]

$$\Delta C(r,z,t) = \left(\frac{d\overline{C}(z)}{dz}\right)_{\text{potential}} \zeta(r,z,t) = GN^2(z)\overline{C}(z)\zeta(r,z,t), \qquad (7)$$

where $G = 1.25$ s^2/m which is empirically determined. Then,

$$\Delta C(r,z,t) = \sum_{j=1}^{j_n} \sum_{k=k_1}^{k_n} \sum_{\theta=\theta_1}^{\theta_n} CIW(j,k,z)\sigma(j,k) \times \text{Re}\{AIW(j,k,\theta) e^{i[k\cos(\theta)r - \omega(j,k)t]}\}, \qquad (8)$$

where

$$CIW(j,k,z) = GN^2(z)\overline{C}(z)\zeta_0 N_\sigma W(j,k,z). \qquad (9)$$

According to the Garrett–Munk spectrum, most of the internal-wave energy is contained over the spatial frequencies $2\times10^{-5}\leqslant k\leqslant 2\times10^{-3}$ rad/m. The corresponding temporal frequencies for these spatial frequencies are $1/24 \leqslant \omega(j,k)-\omega_i \leqslant 2$ cycles/h; thus the ocean is assumed to be frozen. Also, $CIW(j,k,z)$ is approximately constant over k for these spatial frequencies; therefore, $CIW(j,k,z) \simeq CIW(j,z)$ and

$$\Delta C(r,z)=\sum_{j=1}^{j_n}\sum_{k=k_1}^{k_n}\sum_{\theta=\theta_1}^{\theta_n} CIW(j,z)\sigma(j,k) \times \mathrm{Re}\{AIW(j,k,\theta)e^{ik\cos(\theta)r}\}. \quad (10)$$

B. Propagation model—Normal mode with adiabatic approximation

Acoustic pressures are calculated by solving the wave equation using the normal mode method. This method determines the eigenvalues and eigenfunctions that solve the wave equation and satisfy the boundary conditions. The boundary conditions for this deep ocean are a perfectly rigid bottom and a pressure-release surface. For an isotropic point source with a known frequency f [Hz], the result is a Sturm–Liouville problem for which the acoustic wavenumbers, k_m, are all distinct and the acoustic modes, $\Phi_m(z)$, form a complete orthonormal set. KRAKEN, the normal mode program written by Michael B. Porter, is used to compute these acoustic wave numbers and modes.[6,7] Once the wave numbers and modes are computed, the acoustic pressure field generated by a source located at (r_s,z_s) is

$$p(z;r_s,z_s)=p_0\sum_{m=1}^{M}\Phi_m(z_s)\Phi_m(z)\frac{e^{ik_m r_s}}{\sqrt{k_m r_s}}, \quad (11)$$

where z is the depth of the receiver, p_0 is the pressure of the source at $r_s=1$ m, and M is the number of propagating acoustic modes. Normal modes are generally applied to range-independent problems; however, they can also be extended to range-dependent oceans. A range-dependent ocean can be approximated by dividing the range axis into a number of range-independent intervals. The solution can be constructed by using the range-independent normal mode solution and interface conditions (continuity of pressure and normal velocity). We make the adiabatic approximation, which ignores mode coupling (the transfer of energy between modes). After the adiabatic approximation is made, the pressure field becomes

$$p(z;r_s,z_s)=p_0\sum_{m=1}^{M}\Phi_m(z_s,r_s)\Phi_m(z)\frac{e^{i\int_0^{r_s}k_m(r)dr}}{\sqrt{k_m(r_s)r_s}}. \quad (12)$$

Using first-order perturbation theory, the range integrated wave numbers are approximated by

$$\int_0^{r_s}k_m(r)dr\simeq\int_0^{r_s}k_m+\delta k_m(r)dr = k_m r_s+\int_0^{r_s}\delta k_m(r)dr, \quad (13)$$

where[4]

$$\delta k_m(r)=\frac{-(2\pi f)^2}{k_m}\int_0^{\infty}\frac{|\Phi_m(z)|^2}{\overline{C^3}(z)}\Delta C(r,z)dz. \quad (14)$$

After substituting Eq. (10) into Eq. (14) and simplifying,

$$\int_0^{r_s}\delta k_m(r)dr=\sum_{j=1}^{j_n}\Delta k_m(j)G(j), \quad (15)$$

where

$$\Delta k_m(j)=\frac{-(2\pi f)^2}{k_m}\int_0^{\infty}\frac{|\Phi_m(z)|^2}{\overline{C^3}(z)}CIW(j,z)dz \quad (16)$$

and

$$G(j)=\sum_{k=k_1}^{k_n}\sum_{\theta=\theta_1}^{\theta_n}\sigma(j,k) \times \mathrm{Re}\left\{AIW(j,k,\theta)\frac{e^{ik\cos(\theta)r_s}-1}{ik\cos(\theta)}\right\}. \quad (17)$$

C. Processors

For the numerical simulations, the Bartlett processor, optimum uncertain field processor (OUFP), and matched-field processor are implemented to estimate the range of the source. The Bartlett processor assumes the average sound velocity profile is the actual sound velocity profile. This processor is very simple to implement, extremely fast, and quite robust when only small sound speed perturbations are present. The OUFP incorporates environmental uncertainties, i.e., the internal wave, into the signal processing algorithm and optimally estimates the location of the source. The matched-field processor knows the exact realization of the internal-wave field and provides the upper bound on source localization when the environment is known. The localization limits are computed using the probability of correct localization and the rms error of the estimate in high SNR.

1. The optimum uncertain field processor

The optimum uncertain field processor (OUFP), developed by Richardson and Nolte,[2,3] incorporates environmental uncertainty into acoustic source localization by using a Bayesian approach to parameter estimation. Using prior probability distributions of the uncertain source location, propagation, and array parameters, the OUFP calculates $p_{\mathbf{S}|\mathbf{r}}(\mathbf{S}|\mathbf{r})$, the *a posteriori* probability density function of the source location, given the received signal $\mathbf{r}$.

The OUFP assumes the ocean is modeled as a deterministic linear system, allowing the computation of replica fields. It also assumes the signal emitted by the acoustic source, $s(t)=\mathrm{Re}\{A\sqrt{2}e^{i2\pi ft}\}$, is a narrow-band sinusoid with a known frequency f [Hz]. The signal observed at the array of receivers is

$$\mathbf{r}(t)=A\mathbf{s}(\hat{\mathbf{S}},\hat{\mathbf{\Psi}},t)+\mathbf{n}(t), \quad (18)$$

where $\hat{\mathbf{S}}$ is the actual source position and $\hat{\mathbf{\Psi}}$ is the actual ocean environment. The frequency transform of the received signal is of the form

$$\mathbf{P}(\mathbf{r})=A\mathbf{H}(\hat{\mathbf{S}},\hat{\mathbf{\Psi}})+\mathbf{N}, \quad (19)$$

where $\mathbf{H}(\hat{\mathbf{S}},\hat{\mathbf{\Psi}})$ is the acoustic transfer function of the narrow-band source located at $\hat{\mathbf{S}}$ in the ocean $\hat{\mathbf{\Psi}}$. The elements of $\mathbf{P}(\mathbf{r})$ are computed from the discretized time-domain received signal, $r_z[l]$, by

$$P_z=\frac{1}{L}\sum_{l=1}^{L} r_z[l]e^{-i2\pi fTl}, \tag{20}$$

where L is the number of snapshots in time and T is the sampling time. The observation, $\mathbf{P}(\mathbf{r})$, is assumed to contain additive zero-mean complex Gaussian noise, $\mathbf{N}$, with a spatial covariance matrix of $\mathbf{Q}$. In this work the noise is assumed to be isotropic consequently, $\mathbf{Q}=\sigma_N^2\mathbf{I}$. Once the amplitude, A, of the narrow-band source is assumed to be a zero-mean complex Gaussian with variance σ_A^2, the OUFP is determined to be

$$p_{\mathbf{S}|\mathbf{r}}(\mathbf{S}|\mathbf{r})=C(\mathbf{r})p_{\mathbf{S}}(\mathbf{S})\int_{\Psi}\frac{1}{E(\mathbf{S},\mathbf{\Psi})+1}\times\exp\left(\frac{\frac{1}{2}|R(\mathbf{r},\mathbf{S},\mathbf{\Psi})|^2}{E(\mathbf{S},\mathbf{\Psi})+1}\right)p_{\mathbf{\Psi}|\mathbf{S}}(\mathbf{\Psi}|\mathbf{S})\mathbf{d\Psi}, \tag{21}$$

where $C(\mathbf{r})$ is a normalization constant chosen to make $p_{\mathbf{S}|\mathbf{r}}(\mathbf{S}|\mathbf{r})$ a proper probability density function; $\int_{\mathbf{S}}p_{\mathbf{S}|\mathbf{r}}(\mathbf{S}|\mathbf{r})d\mathbf{S}=1$. $\mathbf{S}$ contains the uncertain source location parameters (range and depth). The *a priori* knowledge of the source position, $p_{\mathbf{S}}(\mathbf{S})$, is assumed to be uniform. $\mathbf{\Psi}$ contains the uncertain environmental parameters

$$E(\mathbf{S},\mathbf{\Psi})=\sigma_A^2\mathbf{H}^{\dagger}(\mathbf{S},\mathbf{\Psi})\mathbf{Q}^{-1}\mathbf{H}(\mathbf{S},\mathbf{\Psi})=\frac{\sigma_A^2}{\sigma_N^2}\mathbf{H}^{\dagger}(\mathbf{S},\mathbf{\Psi})\mathbf{H}(\mathbf{S},\mathbf{\Psi}) \tag{22}$$

and

$$R(\mathbf{r},\mathbf{S},\mathbf{\Psi})=\sigma_A\mathbf{H}^{\dagger}(\mathbf{S},\mathbf{\Psi})\mathbf{Q}^{-1}\mathbf{P}(\mathbf{r})=\frac{\sigma_A}{\sigma_N^2}\mathbf{H}^{\dagger}(\mathbf{S},\mathbf{\Psi})\mathbf{P}(\mathbf{r}). \tag{23}$$

At the receiver, the

$$\mathrm{SNR}=E(\hat{\mathbf{S}},\hat{\mathbf{\Psi}})=\frac{\sigma_A^2}{\sigma_N^2}\mathbf{H}^{\dagger}(\hat{\mathbf{S}},\hat{\mathbf{\Psi}})\mathbf{H}(\hat{\mathbf{S}},\hat{\mathbf{\Psi}}). \tag{24}$$

A brute force numerical integration is done when $\mathbf{\Psi}$ contains very few uncertain environmental parameters. When many uncertain environmental parameters are present, the integration over $\mathbf{\Psi}$ can be very efficiently computed using a recently proposed Monte Carlo technique.[3]

2. The wide-band optimum uncertain field processor

The optimum uncertain field processor is expanded to encompass a source containing multiple discrete frequencies. For a wide-band acoustic source Eq. (19) becomes

$$\mathbf{P}_f(\mathbf{r})=A_f\mathbf{H}_f(\hat{\mathbf{S}},\hat{\mathbf{\Psi}})+\mathbf{N}_f,\quad f=f_1,f_2,\ldots,f_F, \tag{25}$$

where $f_1,f_2,\ldots,f_F$ are the discrete frequencies occupied by the source, $\mathbf{H}_f(\hat{\mathbf{S}},\hat{\mathbf{\Psi}})$ is the acoustic transfer function of a source at frequency f located at $\hat{\mathbf{S}}$ in the ocean $\hat{\mathbf{\Psi}}$, A_f is the Fourier transform of the source signal, and $\mathbf{N}_f$ is the additive complex Gaussian noise with a spatial covariance matrix of $\mathbf{Q}_f=\sigma_N^2\mathbf{I}$.

For a random wide-band source, the Fourier transform of the source, A_f, is unknown. However if each amplitude of A_f is assumed to be an independent zero-mean complex Gaussian with variance σ_A^2, the OUFP is

$$p_{\mathbf{S}|\mathbf{r}}(\mathbf{S}|\mathbf{r})=C(\mathbf{r})p_{\mathbf{S}}(\mathbf{S})\int_{\Psi}\prod_{f=f_1}^{f_F}\frac{1}{E_f(\mathbf{S},\mathbf{\Psi})+1}\times\exp\left(\frac{\frac{1}{2}|R_f(\mathbf{r},\mathbf{S},\mathbf{\Psi})|^2}{E_f(\mathbf{S},\mathbf{\Psi})+1}\right)p_{\mathbf{\Psi}|\mathbf{S}}(\mathbf{\Psi}|\mathbf{S})\mathbf{d\Psi}, \tag{26}$$

where

$$E_f(\mathbf{S},\mathbf{\Psi})=\frac{\sigma_A^2}{\sigma_N^2}\mathbf{H}_f^{\dagger}(\mathbf{S},\mathbf{\Psi})\mathbf{H}_f(\mathbf{S},\mathbf{\Psi}) \tag{27}$$

and

$$R_f(\mathbf{r},\mathbf{S},\mathbf{\Psi})=\frac{\sigma_A}{\sigma_N^2}\mathbf{H}_f^{\dagger}(\mathbf{S},\mathbf{\Psi})\mathbf{P}_f(\mathbf{r}). \tag{28}$$

For a coherent wide-band source, the Fourier transform of the source can be written as $A_f=AS_f$, where the complex scalar A contains the uncertainty about the source amplitude and phase over the entire source spectrum and S_f is known. With $S_f=1$, Eq. (19) is

$$\mathbf{P}(\mathbf{r})=A\mathbf{H}(\hat{\mathbf{S}},\hat{\mathbf{\Psi}})+\mathbf{N}, \tag{29}$$

with

$$\mathbf{H}(\hat{\mathbf{S}},\hat{\mathbf{\Psi}})=\begin{bmatrix}\mathbf{H}_{f_1}(\hat{\mathbf{S}},\hat{\mathbf{\Psi}})\\ \mathbf{H}_{f_2}(\hat{\mathbf{S}},\hat{\mathbf{\Psi}})\\ \vdots\\ \mathbf{H}_{f_F}(\hat{\mathbf{S}},\hat{\mathbf{\Psi}})\end{bmatrix}\quad\text{and}\quad\mathbf{N}=\begin{bmatrix}\mathbf{N}_{f_1}\\ \mathbf{N}_{f_2}\\ \vdots\\ \mathbf{N}_{f_F}\end{bmatrix}. \tag{30}$$

When A is assumed to be an zero-mean complex Gaussian with variance σ_A^2, the OUFP for the coherent wide-band source is

$$p_{\mathbf{S}|\mathbf{r}}(\mathbf{S}|\mathbf{r})=C(\mathbf{r})p_{\mathbf{S}}(\mathbf{S})\int_{\Psi}\frac{1}{E(\mathbf{S},\mathbf{\Psi})+1}\times\exp\left(\frac{\frac{1}{2}|R(\mathbf{r},\mathbf{S},\mathbf{\Psi})|^2}{E(\mathbf{S},\mathbf{\Psi})+1}\right)p_{\mathbf{\Psi}|\mathbf{S}}(\mathbf{\Psi}|\mathbf{S})\mathbf{d\Psi}, \tag{31}$$

where

$$E(\mathbf{S},\mathbf{\Psi})=\frac{\sigma_A^2}{\sigma_n^2}\mathbf{H}^{\dagger}(\mathbf{S},\mathbf{\Psi})\mathbf{H}(\mathbf{S},\mathbf{\Psi}) \tag{32}$$

and

$$R(\mathbf{r},\mathbf{S},\mathbf{\Psi})=\frac{\sigma_A}{\sigma_N^2}\mathbf{H}^{\dagger}(\mathbf{S},\mathbf{\Psi})\mathbf{P}(\mathbf{r}). \tag{33}$$

The concatenated $\mathbf{P}(\mathbf{r})$ and $\mathbf{H}(\mathbf{S},\mathbf{\Psi})$ are computed with Eqs. (29) and (30).

For a wide-band source, the total signal-to-noise ratio is defined at the receiver by

$$\mathrm{SNR}=\sum_{f=f_1}^{f_F}\frac{\sigma_A^2}{\sigma_N^2}\,\mathbf{H}_f^{\dagger}(\hat{\mathbf{S}},\hat{\boldsymbol{\Psi}})\mathbf{H}_f(\hat{\mathbf{S}},\hat{\boldsymbol{\Psi}})$$
$$=\frac{\sigma_A^2}{\sigma_N^2}\,\mathbf{H}^{\dagger}(\hat{\mathbf{S}},\hat{\boldsymbol{\Psi}})\mathbf{H}(\hat{\mathbf{S}},\hat{\boldsymbol{\Psi}}). \tag{34}$$

3. The Bartlett processor

The Bartlett processor uses only the mean environment, $\overline{\boldsymbol{\Psi}}$, to localize the source. The Bartlett ambiguity surface is computed by

$$Z_B(\mathbf{S})=\frac{|\mathbf{H}^{\dagger}(\mathbf{S},\overline{\boldsymbol{\Psi}})\mathbf{P}(\mathbf{r})|^2}{\mathbf{H}^{\dagger}(\mathbf{S},\overline{\boldsymbol{\Psi}})\mathbf{H}(\mathbf{S},\overline{\boldsymbol{\Psi}})}. \tag{35}$$

For the coherent wide-band source, the concatenated $\mathbf{P}(\mathbf{r})$ and $\mathbf{H}(\mathbf{S},\hat{\boldsymbol{\Psi}})$ are computed with Eqs. (29) and (30).

4. The matched-field processor

When the actual environment, $\hat{\Psi}$, is completely known Eq. (31) reduces to

$$p_{\mathbf{S}|\mathbf{r}}(\mathbf{S}|\mathbf{r})=C(\mathbf{r})p_S(\mathbf{S})\,\frac{1}{E(\mathbf{S},\hat{\boldsymbol{\Psi}})+1}$$
$$\times\exp\left(\frac{\frac{1}{2}|R(\mathbf{r},\mathbf{S},\hat{\boldsymbol{\Psi}})|^2}{E(\mathbf{S},\hat{\boldsymbol{\Psi}})+1}\right). \tag{36}$$

For the coherent wide-band source, the concatenated $\mathbf{P}(\mathbf{r})$ and $\mathbf{H}(\mathbf{S},\hat{\boldsymbol{\Psi}})$ are computed with Eqs. (29) and (30).

D. The Cramer–Rao lower bound

The Cramer–Rao lower bound (CRLB) can be used to predict the theoretical limit on source localization errors. The CRLB does not provide an algorithm to localize the acoustic source, but rather analytically computes the limits on source range estimates attainable by any processor. These predictions are only valid when the estimation errors are local.

1. The Cramer–Rao lower bound

The Cramer–Rao lower bound (CRLB) can be used to predict the theoretical limit on source localization errors in high SNR. These predictions are only valid when the estimation errors are local, which occurs in the high SNR region only. The CRLB does not provide an algorithm to localize the acoustic source, but rather analytically computes the local limitations on source range and depth estimates attainable by any processor. By incorporating the normal mode solution (along with the adiabatic approximation) to the wave equation and a fully spanning receiver array, Narasimhan and Krolik have computed the CRLB for estimating the range of a narrow-band acoustic source.[11] The Fisher information matrix for source range estimates is

$$\mathbf{J}_{r_s}=\mathrm{SS}\begin{bmatrix}\mathrm{Re}\{\mathbf{k}\}^{\dagger}\mathbf{J}_{\boldsymbol{\Phi\Phi}}\,\mathrm{Re}\{\mathbf{k}\}+\dfrac{\partial\mathbf{h}}{\partial r_s}^{\dagger}\mathbf{J}_{\mathbf{hh}}\dfrac{\partial\mathbf{h}}{\partial r_s} & \mathrm{Re}\{\mathbf{k}\}^{\dagger}\mathbf{J}_{\boldsymbol{\Phi\Phi}}\Delta\mathbf{k}\\ \Delta\mathbf{k}^{\dagger}\mathbf{J}_{\boldsymbol{\Phi\Phi}}\,\mathrm{Re}\{\mathbf{k}\} & \Delta\mathbf{k}^{\dagger}\mathbf{J}_{\boldsymbol{\Phi\Phi}}\Delta\mathbf{k}\end{bmatrix}+\begin{bmatrix}0 & 0\\ 0 & \mathbf{R}_G^{-1}\end{bmatrix}, \tag{37}$$

where SS is the number of independent snapshots of the observation, $\mathbf{k}$ is an $M\times 1$ column vector of the acoustic wave numbers, $\Delta\mathbf{k}$ is an $M\times j_n$ modal phase perturbation matrix defined by Eq. (16),

$$\mathbf{J}_{\boldsymbol{\Phi\Phi}}=\frac{2(\sigma_A^2/\sigma_N^2)^2}{1+(\sigma_A^2/\sigma_N^2)\mathbf{h}^{\dagger}\mathbf{h}}\,(\mathbf{h}^{\dagger}\mathbf{h}\,\mathrm{diag}(\mathbf{h}^2)-\mathbf{h}^2(\mathbf{h}^2)^{\dagger}), \tag{38}$$

$$\mathbf{J}_{\mathbf{hh}}=\frac{2(\sigma_A^2/\sigma_N^2)^2}{1+(\sigma_A^2/\sigma_N^2)\mathbf{h}^{\dagger}\mathbf{h}}\left(\mathbf{h}^{\dagger}\mathbf{h}\mathbf{I}+\frac{1-(\sigma_A^2/\sigma_N^2)\mathbf{h}^{\dagger}\mathbf{h}}{1+(\sigma_A^2/\sigma_N^2)\mathbf{h}^{\dagger}\mathbf{h}}\,\mathbf{h}\mathbf{h}^{\dagger}\right), \tag{39}$$

$$\mathbf{R}_G=\mathrm{diag}(\langle G(j)^2\rangle), \tag{40}$$

$\mathbf{h}$ is an $M\times 1$ column vector of the modal amplitudes whose elements are defined by

$$h_m=\frac{\Phi_m(z_s)e^{-\mathrm{Im}\{k_m\}r_s}}{\sqrt{k_m r_s}}\quad\text{and}\quad h_m^2=\frac{\Phi_m^2(z_s)e^{-2\,\mathrm{Im}\{k_m\}r_s}}{k_m r_s}. \tag{41}$$

The variance of the source range localization error is given by the first diagonal element of $\mathbf{J}_{r_s}^{-1}$. Three CRLBs are computed here: (1) the unperturbed bound for the known ocean; (2) the hybrid bound for the ocean containing an internal wave obeying the Garrett–Munk internal wave spectrum; and (3) the perturbed bound for a completely random ocean. To achieve the unperturbed bound, the exact realization of the internal-wave field must be known making $\langle G(j)^2\rangle=0$ and the range of the source, r_s the only unknown parameter. For the hybrid and perturbed bound, the unknown parameters are the range of the source, r_s, as well as the range integrated internal wave coefficients, $G(j)$ in Eq. (17). The hybrid bound uses the spectrum as *a priori* information on $G(j)$ and computes $\langle G(j)^2\rangle$ from realizations of Eq. (17). The perturbed bound assumes no *a priori* information on the spectrum of the internal wave making $\langle G(j)^2\rangle=\infty$.

2. The wide-band Cramer–Rao lower bound

The CRLB is extended to provide the theoretical limit on range estimates using either a random or a coherent wide-band source. For the random wide-band source, the Fisher information matrix for source range estimates is

$$\mathbf{J}_{r_s}=\sum_{f=f_1}^{f_F}\begin{bmatrix}\mathrm{Re}\{\mathbf{k}_f\}^{\dagger}\mathbf{J}_{\boldsymbol{\Phi\Phi}}\,\mathrm{Re}\{\mathbf{k}_f\}+\dfrac{\partial\mathbf{h}_f}{\partial r_s}^{\dagger}\mathbf{J}_{\mathbf{hh}}\dfrac{\partial\mathbf{h}_f}{\partial r_s} & \mathrm{Re}\{\mathbf{k}_f\}^{\dagger}\mathbf{J}_{\boldsymbol{\Phi\Phi}}\Delta\mathbf{k}_f\\ \Delta\mathbf{k}_f^{\dagger}\mathbf{J}_{\boldsymbol{\Phi\Phi}}\,\mathrm{Re}\{\mathbf{k}_f\} & \Delta\mathbf{k}_f^{\dagger}\mathbf{J}_{\boldsymbol{\Phi\Phi}}\Delta\mathbf{k}_f\end{bmatrix}+\begin{bmatrix}0 & 0\\ 0 & \mathbf{R}_G^{-1}\end{bmatrix}, \tag{42}$$

where $\mathbf{k}_f$ is the $M\times 1$ column vector of the acoustic wave numbers for the source at frequency f, $\mathbf{\Delta k}_f$ is the corresponding $M\times j_n$ modal phase perturbation matrix defined by Eq. (16), and $\mathbf{h}_f$ is an $M\times 1$ column vector of the modal amplitudes computed by Eq. (41). $\mathbf{J_{\Phi\Phi}}$ and $\mathbf{J_{hh}}$ are computed by Eqs. (38) and (39) with $\mathbf{h}=\mathbf{h}_f$.

For the coherent wide-band source the Fisher information matrix for source range estimates is

$$\mathbf{J}_{r_s}=\begin{bmatrix} \mathrm{Re}\{\mathbf{k}\}^{\dagger}\mathbf{J_{\Phi\Phi}}\,\mathrm{Re}\{\mathbf{k}\}+\dfrac{\partial \mathbf{h}}{\partial r_s}^{\dagger}\mathbf{J_{hh}}\dfrac{\partial \mathbf{h}}{\partial r_s} & \mathrm{Re}\{\mathbf{k}\}^{\dagger}\mathbf{J_{\Phi\Phi}}\mathbf{\Delta k} \\ \mathbf{\Delta k}^{\dagger}\mathbf{J_{\Phi\Phi}}\,\mathrm{Re}\{\mathbf{k}\} & \mathbf{\Delta k}^{\dagger}\mathbf{J_{\Phi\Phi}}\mathbf{\Delta k} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{R}_G^{-1} \end{bmatrix}, \tag{43}$$

with

$$\mathbf{k}=\begin{bmatrix} \mathbf{k}_{f_1} \\ \mathbf{k}_{f_2} \\ \vdots \\ \mathbf{k}_{f_F} \end{bmatrix}, \quad \mathbf{\Delta k}=\begin{bmatrix} \mathbf{\Delta k}_{f_1} \\ \mathbf{\Delta k}_{f_2} \\ \vdots \\ \mathbf{\Delta k}_{f_F} \end{bmatrix}, \quad \text{and} \quad \mathbf{h}=\begin{bmatrix} \mathbf{h}_{f_1} \\ \mathbf{h}_{f_2} \\ \vdots \\ \mathbf{h}_{f_F} \end{bmatrix}. \tag{44}$$

$\mathbf{J_{\Phi\Phi}}$ and $\mathbf{J_{hh}}$ are computed by Eqs. (38) and (39) with the concatenated $\mathbf{h}$ from Eq. (44).

II. SIMULATIONS AND RESULTS

For all analytical and numerical simulations, an acoustic source with a specified spectrum is placed 1000 km in range ($r_s=1000$ km) and 800 m in depth ($z_s=800$ m), which is on the channel axis. This source is placed in an ocean containing a known range-independent sound-speed profile, $C(z)$, with a random range-dependent sound-speed perturbation, $\Delta C(r,z)$, due to an internal wave. Realizations of the internal wave are computed using 25 ($j_n=25$) discrete modes. Each discrete internal-wave mode propagates at 15 known spatial frequencies spaced logarithmically between 2×10^{-5} and 2×10^{-3} rad/m and at 25 uniformly spaced random angles. The complex amplitudes, $AIW(j,k,\theta)$, of each discrete internal-wave mode at each spatial frequency and propagating angle are random. The sound-speed perturbations, $\Delta C(r,z)$, along with the corresponding rms levels, $\langle \Delta C(z)\rangle$ rms, caused by a realization of this random internal wave are in Fig. 2. The acoustic wave propagation of this source in this ocean is simulated by the normal mode solution to the wave equation using KRAKEN and the adiabatic approximation.[6,7] The acoustic pressure field is sampled with a fully spanning array of receivers to allow modal filtering. First, purely analytical limitations on source range estimates are computed using the CRLB in Sec. II A. Because the CRLB does not provide the algorithm to localize the acoustic source, they only compute potential source localization limits. The actual source localization limits are computed with the optimal signal processing algorithm. Ambiguity plots and processor performance curves are computed in Sec. II B.

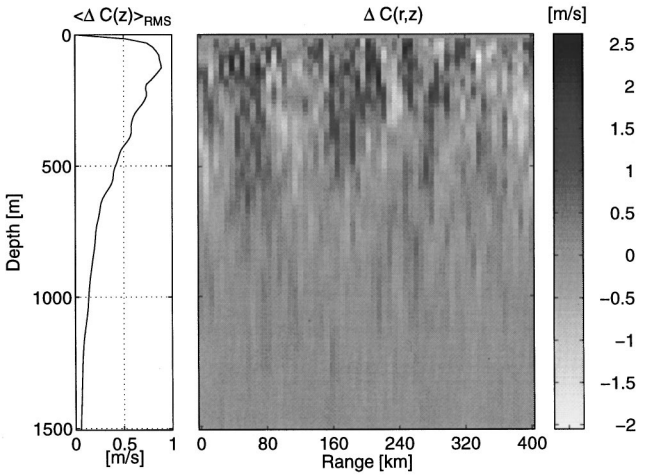


FIG. 2. Sound-speed perturbations, $\Delta C(r,z)$, and the corresponding $\langle \Delta C(r,z)\rangle_{\mathrm{rms}}$ caused by a realization of a very random internal wave.

A. Analytical results

The Cramer–Rao lower bound is used to predict the analytical limits on source range estimates without the computational burden of implementing any signal processing algorithm. The CRLB computes potential limits of source range estimates for wide-band acoustic sources with various source spectrum *a priori*.

As the spectrum of the acoustic source spans more frequencies, the accuracy of range estimates should improve with this increase in signal information. In these analytical computations, the acoustic source contains two discrete frequencies, $f_1=25$ Hz and f_2 ranges from 25 to 125 Hz. The adiabatic assumption is not expected to be valid for an acoustic source at 125 Hz, but it is computed for illustration purposes and provides the analytical limits on source range estimates for acoustic sources with higher frequencies.[1] The hybrid and unperturbed CRLBs for the random wide-band source are computed with Eq. (42) in Fig. 3. The $f_1=25$ Hz and $f_2=25$ Hz case is equivalent to having two

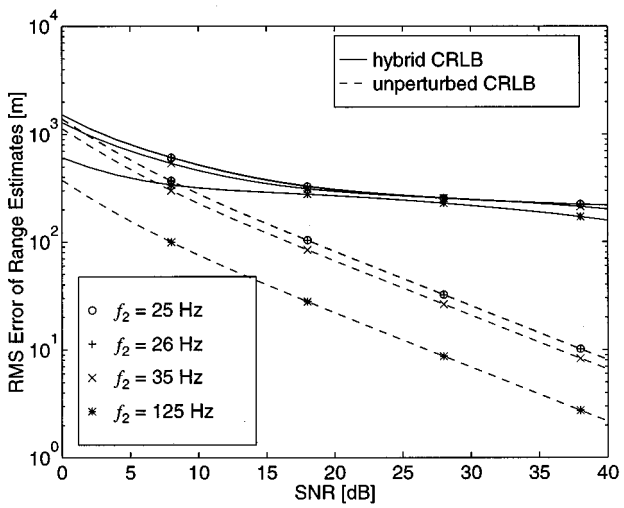


FIG. 3. The hybrid and unperturbed CRLBs for the random wideband source containing two discrete frequencies, $f_1=25$ Hz and f_2 ranging from 25 to 125 Hz.

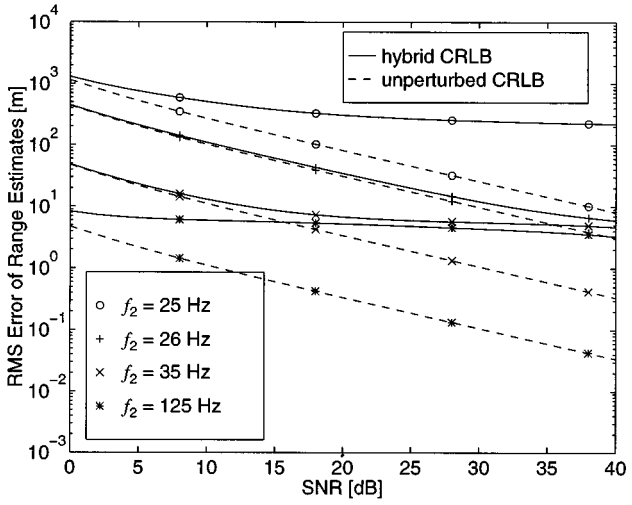


FIG. 4. The hybrid and unperturbed CRLBs for the coherent wideband source containing two discrete frequencies, $f_1 = 25$ Hz and f_2 ranging from 25 to 125 Hz.

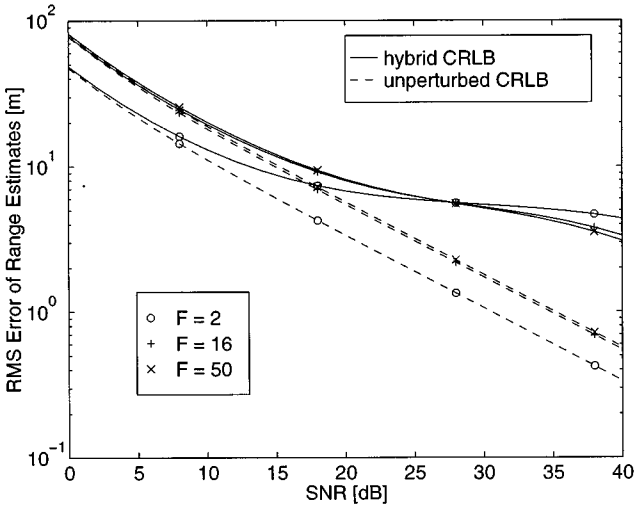


FIG. 5. The hybrid and unperturbed CRLBs for a coherent wideband acoustic source containing 2, 16, and 50 discrete frequencies uniformly spaced from 20 to 30 Hz.

independent snapshots of a 25 Hz narrow-band source. According to the hybrid CRLB for this random internal-wave ocean, range estimates do not benefit from a bandwidth increase for a random wide-band source. The OUFP achieves these predicted limits using a narrow-band source.[5] Figure 4 contains the hybrid and unperturbed CRLBs computed for the coherent wide-band source using Eq. (43). The $f_1 = 25$ Hz and $f_2 = 25$ Hz case is now equivalent to having two coherent snapshots of a 25-Hz narrow-band source. With the coherent wide-band source in this random internal-wave ocean, the hybrid CRLB significantly improves with frequency separation. Once the frequency separation reaches 10 Hz for the 25-Hz source, this improvement becomes less significant. Thus the wide-band acoustic source must be coherent in order to achieve any significant improvement in range estimates. Baggeroer and Schmidt also found that a coherent wide-band source performed significantly better than a random wide-band source when using the Cramer–Rao lower bounds to estimate temperature perturbations.[12]

Intuitively, range estimates should improve as more discrete frequencies are present in a wide-band acoustic source with a constant bandwidth. The hybrid and unperturbed CRLBs are computed for a coherent wide-band acoustic source containing $F = 2$, 16, and 50 discrete frequencies uniformly spaced from 20 to 30 Hz in Fig. 5. Please note that the SNR is normalized over the number of frequencies by Eq. (34). Also for the $F = 2$ case only two discrete frequencies, $f_1 = 20$ Hz and $f_2 = 30$ Hz, are present and source does not have a continuous bandwidth. As evident from the hybrid CRLB for this random internal-wave ocean, range estimates do not significantly improve as the coherent wide-band source contains more discrete frequencies.

B. Numerical results

The Cramer–Rao lower bound provides analytical insight into the limits on source range estimates, but the optimal signal processing algorithm is necessary to compute the actual limits on source localization. Using matched-field processor ambiguity plots, the effects of having a coherent wide-band source are investigated. Finally, the limits of source range estimates are computed using many ambiguity plots from the OUFP.

The sensitivity of bandwidth on range estimates are illustrated with ambiguity plots computed by the matched-field processor. Figure 6 contains a typical ambiguity plot for a 25-Hz narrow-band acoustic source using the matched-field processor at an SNR of 40 dB. The ambiguity plot computed by the matched-field processor for a coherent wide-band source with two discrete frequencies, 20 and 30 Hz, at an SNR of 40 dB is computed in Fig. 7. With the two discrete coherent frequencies at 20 and 30 Hz, the peak of the true location is much sharper than with the narrow-band source; however, the ambiguity plot of the coherent source contains many sharp peaks. The narrow-band ambiguity plot almost appears to be the envelope of the ambiguity plot computed with the coherent wide-band source. To explain this appear-

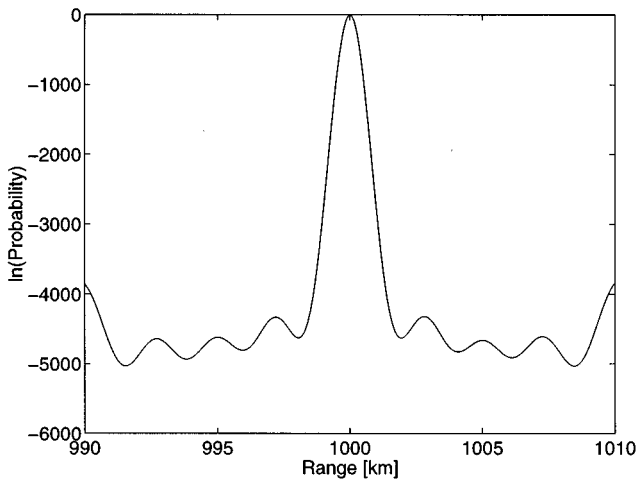


FIG. 6. Typical matched-field processor ambiguity plot for a 25-Hz narrow-band acoustic source at an SNR of 40 dB.

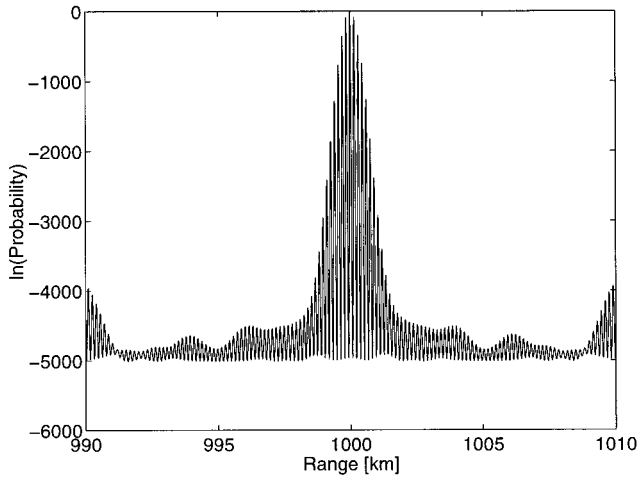


FIG. 7. Typical matched-field processor ambiguity plot for a source containing two coherent frequencies, $f_1=20$ Hz and $f_2=30$ Hz, at an SNR of 40 dB.

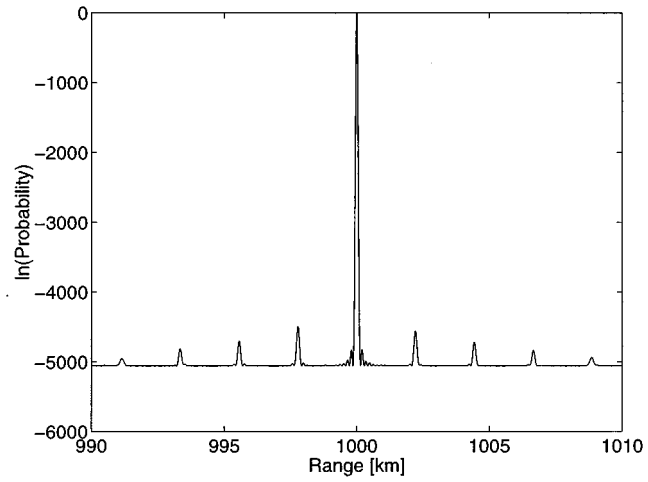


FIG. 8. Typical matched-field processor ambiguity plot for a source containing 16 coherent frequencies spaced uniformly between 20 and 30 Hz at an SNR of 40 dB.

ance, the normal mode solution to the wave equation and the matched-field processor are investigated. In the normal mode solution, if the number of acoustic modes, M, is approximately constant and the most significant dependence on source frequency is the acoustic wave numbers, $\mathbf{k}_f$, then with two coherent frequencies ($F=2$) and Eq. (11),

$$\mathbf{H}(\mathbf{S},\hat{\boldsymbol{\Psi}})=\begin{bmatrix}\mathbf{H}_{f_1}(\mathbf{S},\hat{\boldsymbol{\Psi}})\\ \mathbf{H}_{f_2}(\mathbf{S},\hat{\boldsymbol{\Psi}})\end{bmatrix}\simeq\begin{bmatrix}\mathbf{H}_{\overline{f}}(\mathbf{S},\hat{\boldsymbol{\Psi}})e^{-i(k_{\Delta f}/2)r}\\ \mathbf{H}_{\overline{f}}(\mathbf{S},\hat{\boldsymbol{\Psi}})e^{i(k_{\Delta f}/2)r}\end{bmatrix}, \tag{45}$$

and, under high SNR,

$$\mathbf{P}(\mathbf{r})=\begin{bmatrix}\mathbf{P}_{f_1}(\mathbf{r})\\ \mathbf{P}_{f_2}(\mathbf{r})\end{bmatrix}\simeq\begin{bmatrix}\mathbf{P}_{\overline{f}}(\mathbf{r})e^{-i(k_{\Delta f}/2)r_s}\\ \mathbf{P}_{\overline{f}}(\mathbf{r})e^{i(k_{\Delta f}/2)r_s}\end{bmatrix}, \tag{46}$$

with the center frequency, $\overline{f}$, and $\mathbf{k}_{f_F}-\mathbf{k}_{f_1}\simeq k_{\Delta f}\mathbf{1}$, where $\mathbf{1}$ is an $M\times 1$ column vector. The exact internal-wave realization is known with the matched-field processor and for high SNR, Eq. (36) is approximated by

$$\ln(p_{\mathbf{S}|\mathbf{r}}(\mathbf{S}|\mathbf{r}))\simeq\frac{|\mathbf{H}^{\dagger}(\mathbf{S},\hat{\boldsymbol{\Psi}})\mathbf{P}(\mathbf{r})|^2}{\mathbf{H}^{\dagger}(\mathbf{S},\hat{\boldsymbol{\Psi}})\mathbf{H}(\mathbf{S},\hat{\boldsymbol{\Psi}})}, \tag{47}$$

which is the Bartlett processor in a known environment. After substituting $\mathbf{P}(\mathbf{r})$ and $\mathbf{H}(\mathbf{S},\hat{\boldsymbol{\Psi}})$ from Eqs. (46) and (45) into Eq. (47),

$$\ln(p_{\mathbf{S}|\mathbf{r}}(\mathbf{S}|\mathbf{r}))\simeq\frac{|\mathbf{H}^{\dagger}_{\overline{f}}(\mathbf{S},\hat{\boldsymbol{\Psi}})\mathbf{P}_{\overline{f}}(\mathbf{r})[e^{-i(k_{\Delta f}/2)(r_s-r)}+e^{i(k_{\Delta f}/2)(r_s-r)}]|^2}{\mathbf{H}^{\dagger}_{\overline{f}}(\mathbf{S},\hat{\boldsymbol{\Psi}})\mathbf{H}_{\overline{f}}(\mathbf{S},\hat{\boldsymbol{\Psi}})2} \tag{48}$$

or

$$\ln(p_{\mathbf{S}|\mathbf{r}}(\mathbf{S}|\mathbf{r}))\simeq\frac{|\mathbf{H}^{\dagger}_{\overline{f}}(\mathbf{S},\hat{\boldsymbol{\Psi}})\mathbf{P}_{\overline{f}}(\mathbf{r})|^2}{\mathbf{H}^{\dagger}_{\overline{f}}(\mathbf{S},\hat{\boldsymbol{\Psi}})\mathbf{H}_{\overline{f}}(\mathbf{S},\hat{\boldsymbol{\Psi}})}2\cos^2\left(\frac{k_{\Delta f}}{2}(r_s-r)\right), \tag{49}$$

which is the narrow-band ambiguity function multiplied by a squared cosine. This is just an approximation because in this example the number of acoustic modes does not remain constant from 20 to 30 Hz and, therefore, the envelope of the coherent ambiguity plot does not exactly match the narrow-band ambiguity plot. Because the ambiguity plot is multimodal, simply increasing the bandwidth may not improve source range estimates.

To illustrate the effect of frequency density of a coherent wide-band source, another ambiguity plot is computed using the matched field processor. This wide-band source has the same coherent bandwidth as in Fig. 7, but with 16 uniformly spaced frequencies between 20 and 30 Hz. The matched-field processor ambiguity plot for this coherent wideband source at an SNR of 40 dB is computed in Fig. 8. With more coherent frequencies, the true peak becomes very sharp and little ambiguity remains. With F coherent frequencies present in the source, Eq. (48) becomes

$$\ln(p_{\mathbf{S}|\mathbf{r}}(\mathbf{S}|\mathbf{r}))\simeq\frac{|\mathbf{H}^{\dagger}_{\overline{f}}(\mathbf{S},\hat{\boldsymbol{\Psi}})\mathbf{P}_{\overline{f}}(\mathbf{r})\Sigma_{f=1}^{F}e^{i[(f-1)/(F-1)-1/2]k_{\Delta f}(r_s-r)}|^2}{\mathbf{H}^{\dagger}_{\overline{f}}(\mathbf{S},\hat{\boldsymbol{\Psi}})\mathbf{H}_{\overline{f}}(\mathbf{S},\hat{\boldsymbol{\Psi}})F}, \tag{50}$$

which can also be expressed as the narrow-band ambiguity function multiplied by cosines at different frequencies. Thus ambiguity is resolved with more coherent frequencies. This phenomena is not predicted by the CRLBs in Sec. II A.

Many ambiguity plots are computed when determining the actual limits on source range estimates. The Bartlett processor, OUFP, and matched-field processor are implemented to estimate the range of the source in this internal-wave environment. Each processor searches between 990 and 1110 km in 1-m increments. All processors use the complex modal coefficients to localize the source. The location of the source is estimated to be at the maximum of the ambiguity plot. A range estimate is considered correct if it is within 100 m of the actual range (i.e., 1000 km$\pm$100 m). Fifty independent internal waves are simulated in computing the performance curves for each processor. The performance curves are com-

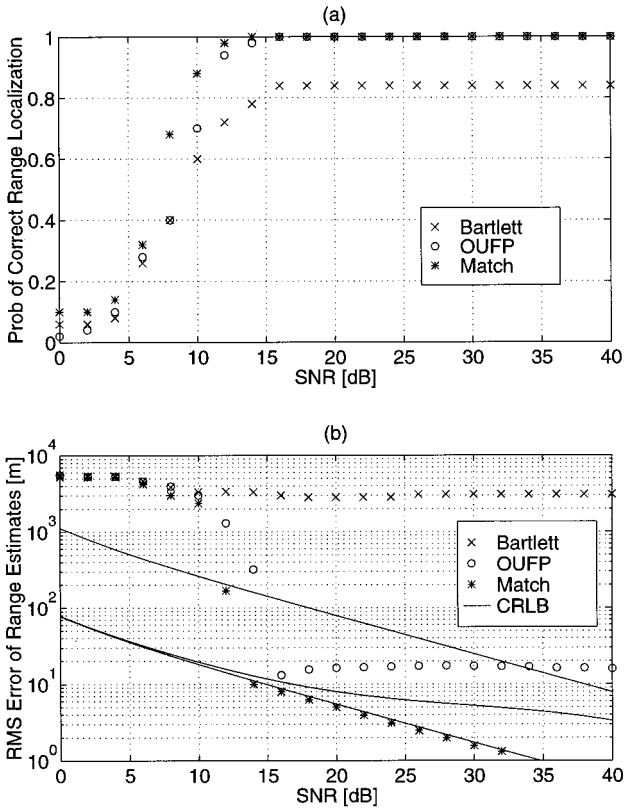


FIG. 9. Processor performance in very random internal-wave environments shown as: (a) probability of correct range localization, and (b) rms error of range estimates.

puted for SNRs from 0 to 40 dB. The SNR of each observation is set by fixing the amplitude of the source to σ_A and adding the zero-mean complex Gaussian noise with variance σ_N^2. The OUFP uses 100 internal-wave realizations in the Monte Carlo numerical integration.

The performance of each processor in this internal-wave environment is shown in Fig. 9(a) as the probability of correct range localization and in Fig. 9(b) as the rms error on range estimates. The unperturbed CRLB, the hybrid CRLB, and the perturbed CRLB are also shown in Fig. 9(b). For low SNR, the rms error of range estimates for all processors is a function of the search window, which should be infinite, since the errors are no longer local. The CRLB only predicts the local error associated with the source localization; however, under low SNR, global errors are present. This thresholding phenomena occurs at SNR$\simeq$15 dB. The Bartlett processor uses the source spectrum *a priori* optimally, but does not achieve the performance levels predicted by the hybrid CRLB. The OUFP attains the same performance level predicted by the hybrid CRLB. A very accurate integration over the environment $\mathbf{\Psi}$ is necessary for the performance of the OUFP to exactly achieve the accuracy predicted by the hybrid CRLB. The hybrid CRLB theoretically predicts the accuracy of range estimates and the OUFP is the signal processing algorithm that achieves this performance level. The performance of the matched-field processor, which has complete knowledge of each internal-wave realization, lies on the unperturbed CRLB. Using a coherent wide-band source with many discrete frequencies spanning the bandwidth, significant gains in range estimates are achieved by optimally incorporating this signal spectrum into the localization algorithm *a priori*.

III. CONCLUSIONS

By optimally incorporating internal-wave characteristics and source spectrum into the OUFP, the limits on source range estimates of a wide-band acoustic source in the presence of internal waves have been computed. The CRLB provides valuable analytical insight into the limits on source range estimates for various source spectrum without the computational burden of implementing an algorithm. According to the hybrid CRLB, range estimates do not benefit from a increase in bandwidth of a random source in an ocean containing an internal wave. Significant gains are made in range estimates when using a coherent wide-band acoustic source with a frequency separation of at least 10 Hz on the 25-Hz source.

The CRLB provides an effective analytical prediction of source localization limitations at high SNR, but does not predict global localization errors which occur in low SNR conditions; therefore, the optimal signal processing algorithm is necessary to compute the actual limits on source localization. When localizing a coherent wide-band source using the matched-field processor, the ambiguity plot contains many sharp peaks. As the coherent wide-band source contains more discrete frequencies, ambiguity is resolved. The limits of source range estimates are computed using many ambiguity plots from the OUFP. The Bartlett processor incorporates the source spectrum optimally and consequently performs quite well; however, under high SNR, it does not achieve the performance levels predicted by the hybrid CRLB. The OUFP attains the performance level predicted by the hybrid CRLB; however, a very accurate integration over the environment is necessary for the OUFP to exactly achieve the accuracy predicted by the hybrid CRLB. The hybrid CRLB theoretically predicts the accuracy of range estimates and the OUFP is the signal processing algorithm that achieves this performance bound. By knowing the ocean environment, the matched-field processor computes the upper bound on performance and, in high SNR conditions, performs along the theoretical unperturbed CRLB. Using a coherent wide-band source with many discrete frequencies spanning the bandwidth, significant gains in range estimates are achieved by optimally incorporating this signal spectrum into the localization algorithm *a priori*.

ACKNOWLEDGMENT

Support for this work has been provided by the Office of Naval Research (Ocean Acoustics, code 321OA).

[1] J. A. Colosi and S. M. Flatte, ''Mode coupling by internal waves for the multimegameter acoustic propagation in the ocean,'' J. Acoust. Soc. Am. **100**, 3607–3620 (1996).
[2] A. M. Richardson and L. W. Nolte, ''*A posteriori* probability source localization in an uncertain sound speed, deep ocean environment,'' J. Acoust. Soc. Am. **89**, 2280–2284 (1991).
[3] J. A. Shorey, L. W. Nolte, and J. L. Krolik, ''Computationally efficient monte carlo estimation algorithms for matched field processing in uncertain ocean environments,'' J. Comp. Acoust. **2**, 285–314 (1994).

[4] J. L. Krolik, ''Matched-field minimum variance beamforming in a random ocean channel,'' J. Acoust. Soc. Am. **92**, 1408–1419 (1992).
[5] P. J. Book and L. W. Nolte, ''Narrow-band source localization in the presence of internal waves for 1000-km range and 25-Hz acoustic frequency,'' J. Acoust. Soc. Am. **101**, 1336–1346 (1997).
[6] M. B. Porter and E. L. Reiss, ''A numerical method for ocean acoustic normal modes,'' J. Acoust. Soc. Am. **76**, 244–252 (1984).
[7] F. Jensen, W. Kuperman, M. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994).
[8] S. Levitus, *Climatological Atlas of the World Ocean* (National Oceanic and Atmospheric Administration, Rockville, MD, 1982).
[9] J. Colosi, S. Flatte, and C. Bracher, ''Internal-wave effects on 1000 km oceanic acoustic pulse propagation: Simulation and comparison with experiment,'' J. Acoust. Soc. Am. **84**, 452–467 (1994).
[10] S. M. Flatte, R. Dashen, W. Munk, K. Watson, and F. Zacariasen, *Sound Transmission Through a Fluctuating Ocean* (Cambridge U.P., Cambridge, 1979).
[11] S. Narasimhan and J. L. Krolik, ''Fundamental limits on acoustic source range estimation performance in uncertain ocean channels,'' J. Acoust. Soc. Am. **97**, 215–226 (1995).
[12] A. B. Baggeroer and H. Schmidt, ''Cramer–Rao bounds for matched field tomography and ocean acoustic tomography,'' ICASSP, 2763–2766 (1995).

Sound-speed measurements in the surface-wave layer

Eric Terrill and W. Kendall Melville
Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0230

(Received 28 January 1997; revised 21 July 1997; accepted 23 July 1997)

Wave breaking at the surface of the ocean entrains bubbles, significantly modifying the phase speed and attenuation of acoustic waves propagating through the resulting two-phase medium. An autonomous buoy system was developed that directly measures sound speed at 3.33, 5, and 10 kHz at seven depths ranging from 0.7 to 7 m through the use of a travel-time technique. Simultaneous measurements at each depth are obtained at a 2-Hz rate, allowing observation of the unsteady sound-speed field from individual bubble injection events, as well as the calculation of mean sound speeds. The travel-time technique allows a direct measurement of the sound speed, eliminating the uncertainties common with inferring sound speeds from bubble population data. The sound speed buoy was deployed in the North Atlantic during the winter of 1993–94 as part of the Acoustic Surface Reverberation Experiment (ASREX). Our aim was to characterize the highly variable near-surface sound-speed field under varying environmental conditions. Forty-three days of data were obtained spanning several storm cycles with wind speeds and significant wave heights reaching 20 m/s and 8 m, respectively. During periods of intense wave breaking, average sound speeds below 1000 m/s were observed at the 0.7-m measurement depth while instantaneous sound speeds during individual events approached values as low as 300 m/s. Furthermore, the data suggest that the dispersive effects of bubbles may extend to frequencies as low as 5 kHz near the surface during storms. Strong correlations of the mean and rms sound speed with the overlying wind and wave fields were found. © *1997 Acoustical Society of America.* [S0001-4966(97)02711-2]

PACS numbers: 43.30.Es, 43.30.Nb, 43.30.Xm [SAC-B]

INTRODUCTION

The study of breaking waves and the associated bubble injection has received considerable attention from acousticians and oceanographers interested in air–sea interaction. During periods of rough weather, the scales of wave breaking tend to increase with increasing sea states, vigorously mixing the surface waters and injecting bubbles into the water column. Observations of bubbles using a variety of techniques show their presence to depths of O(10) m while the bulk of the entrained air resides in the first few meters below the surface. Bubble clouds (or plumes) form as a result of wave breaking, with near-surface turbulence and Langmuir circulations contributing to their evolution.[1,2] Wave breaking also plays a primary role in the dissipation of energy in the wave field, transferring energy and momentum to surface currents and turbulence. This turbulence is thought to be the main source of mixing in the near-surface waters, enhancing both heat and gas transfer across the interface.

Interest in the acoustical properties of entrained bubbles has existed for a number of years, stemming from the introduction of sonar systems in World War II. Theory and observations show that bubbles can be both efficient scatters and absorbers of sound. For example, for acoustic frequencies at or near its lowest resonant frequency, a bubble has a scattering cross section $O(1000)$ times that of a solid particle of similar dimensions. Furthermore, both individual resonances and collective oscillations of groups of bubbles are known to contribute significantly to the ambient noise spectrum of the ocean.[3–7] One area of recent interest has been the effect of near-surface bubbles on low-frequency scattering and reverberation from the air–sea interface. High levels of low-frequency scatter during high wind speeds suggest that entrained bubbles near the surface change the nature of the backscatter.[8,9]

Methods used in the past to measure bubbles and their distributions have included both optical and acoustical techniques. Optical techniques pioneered by Johnson and Cooke[10] and later used by others[11–13] are plagued by small sampling volumes, especially for large bubbles, and the failure to identify small bubbles. However, a recent development of an optical ''depth of focus'' technique offers promise of resolving bubble radii ranging from 20 to 1000 μm.[14] The techniques used have been slow in their sampling rates $O(0.1\ \text{Hz})$ or of short duration $O(30\ \text{min})$, and most require laborious processing (in some cases, manual counting). Despite the shortcomings of past optical data, the findings of Johnson and Cooke[10] of a slope in the bubble size distribution of a^{-3} to $a^{-4.5}$ (for bubble radii a greater than 60 μm), has been incorporated into models of gas transfer and acoustic propagation.[1,15] Due to data storage and processing constraints, optical techniques are currently unsuitable for a remotely operated system which would be used for obtaining long time series of $O(1\ \text{month})$ duration.

Acoustic methods of measuring bubbles are attractive because of the ability to design robust equipment for the harsh operating conditions of the surface layer and the existence of digital signal processing techniques designed for large amounts of data. Both *in-situ* and remote acoustic methods have been used in past bubble studies. The *in-situ* methods initiated by Medwin and his colleagues[16–19] used a variety of techniques to study bubbles and their affects on acoustics in calm conditions. Studying the scattering and ab-

sorption of sound pulses over a range of frequencies, they inferred bubble populations over time-averaged signals $O(10$ min) in waters 3 m and deeper. Later work included the measurements of dispersive and nondispersive sound speeds. Sound speeds were calculated by measuring phase shifts in a continuous-wave signal between two fixed hydrophones. Again, this work was done in relatively deep water and under benign conditions. More recently, Medwin and his colleagues have developed a bubble resonator that measures the response of bubbles ensonified in the resonant chamber.[19] A drawback of this device was the long sampling period of 10 s. An excellent review of the performance of this device in the laboratory is given by Su.[20]

Efforts to remotely measure bubbles in the surface layer have been made by Vagle and Farmer.[21] Backscatter data from multifrequency, upward looking sonars was interpreted to provide bubble size distributions for radii $\sim 10-100$ μm. Bubble size distribution slopes were reported to vary from -1 to -8 during 45 min of sampling. In earlier work, Farmer and Vagle[22] integrated over the measured bubble populations to arrive at a void fraction, defined as $\beta = (4/3)\pi \int a^3 n(a) da$ [typically $n(a)$ is defined as the number of bubbles per m^3 per micron radii increment]. Using known relationships between low-frequency sound speed and void fraction of air in water,[23,24] a depth profile of the sound-speed anomaly due to air injection from breaking waves was calculated. For winds of 12 m/s during the FASINEX experiment, a sound-speed anomaly of 19.0 m/s ($\beta \sim 10^{-6}$) at the surface was reported.

More recent efforts to directly measure near-surface sound speeds were undertaken in a series of field measurements by Lamarre and Melville.[25,26] Their direct measurements of sound speed eliminates the difficulties involved with inferring sound speeds from bubble populations or void fractions. Lamarre and Melville[25] pointed out that uncertainties in bubble population data, especially for large bubbles, can lead to large errors when interpreting upward looking sonar data for sound-speed profiles. The integrals used for calculating void fraction are sensitive to both the slopes of the power laws that describe the bubble size distributions and the limits of bubble sizes over which the integration is performed.

At acoustic frequencies much less than the resonant frequencies of the bubbles in the air/seawater mixture, the speed of sound is $c_m = 1/(\rho_m K_m)^{1/2}$, where the density and compressibility of the mixture are given by $\rho_m = \beta \rho_a + (1 - \beta)\rho_w$ and $K_m = \beta K_a + (1-\beta)K_w$, respectively. Since the density and compressibility of the mixture depend upon the amount of air entrained, the sound speed is a function of the void fraction of air. This low-frequency limit of the sound speed is given by Wood's curve shown in Fig. 1 and has been shown to be valid for sound speeds as low as 30 m/s.[27] It is clear that the sound speed is sensitive to void fraction with the reduction providing an excellent indication of entrained air. For example, at void fractions of $O(10^{-4})$ the sound speed is roughly halved.

The instrument developed by Lamarre and Melville measured travel times of acoustic pulses transmitted across a horizontal pathlength of approximately 400 mm. Using a vertical array of instruments, simultaneous measurements of sound speed were made at several depths, starting at 0.5 m and extending to 4 m, at sampling rates of 4 Hz. The high temporal resolution of the system allows observations of the changing sound-speed profile due to wave breaking.

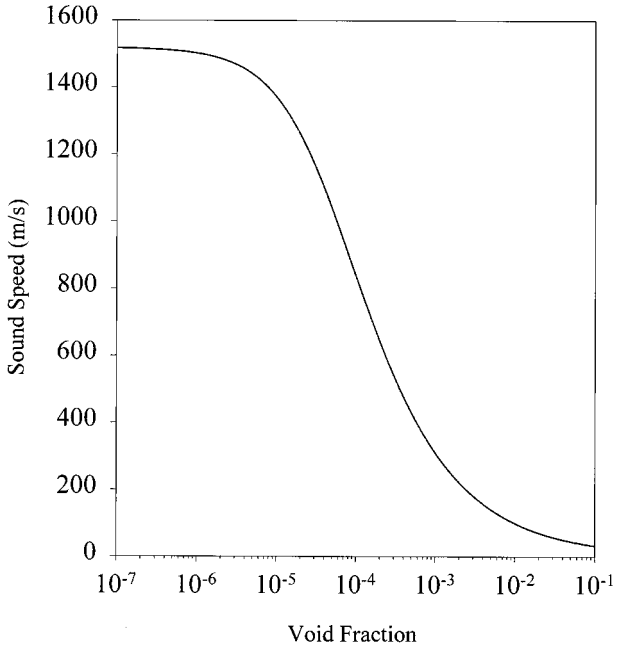


FIG. 1. Wood's curve showing sound speed as a function of void fraction.

Experiments were conducted off San Diego, California and in Buzzard's Bay, Massachusetts. For both sites, winds never exceeded 8 m/s. During the experiments, the instrument was tethered to the support ship where data acquisition and control systems were located. At a depth of 0.5 m, observations of recently generated bubble plumes produced reductions of $O(800$ m/s) while several reductions of $O(100$ m/s) were frequently measured during moderate wind conditions. Averages of the sound-speed reduction time series were computed for the various depths to investigate the depth dependence of the sound-speed field. For winds of 7 to 8 m/s, an exponential decrease with depth was observed in qualitative agreement with the sound-speed profiles inferred by Farmer and Vagle;[22] however, significant differences were found in the constants used to describe the exponential fit. Attempts at parametrizing the average sound-speed reductions with the wind speed offered some correlation but significant scatter in the data suggested the need for a larger data set.

Lamarre and Melville extended the measurements to include simultaneous sound-speed measurements at multiple frequencies from 5 to 40 kHz. A broadband pulse was bandpass filtered at various frequency bands, allowing travel-time and attenuation measurements to be made at the various bands. In the 8-m/s winds, the sound speed was observed to be nondispersive at frequencies below approximately 20 kHz.

In Sec. I of this paper, an autonomous buoy system designed for long-term sound-speed measurements in the surface wave zone is described. Section II describes the ASREX field experiment. Section III summarizes the quality of the

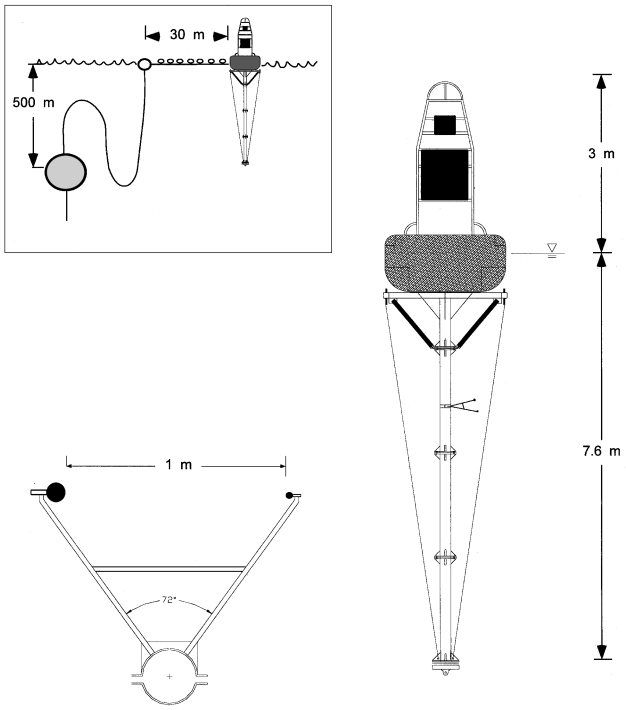


FIG. 2. Schematic of the sound-speed buoy system. Insets include the inverse catenary of the deep water mooring and a plan view of the sound-speed modules which affix to the spar of the buoy.

data and establishes error bounds on the measurements. Section IV consists of analyses of the mean and rms sound speeds in the context of the environmental parameters measured during the experiment. In Sec. V the fine scale properties of the sound-speed field are examined, providing insight into the variability of the near-surface layer. Section VI concludes the paper with a discussion of the results obtained from the analysis.

I. EQUIPMENT DESCRIPTION

A buoy was designed to support the acoustic instrumentation under a wide range of sea states. The resulting structure has an overall length of 10.7 m with a maximum diameter of 2.44 m at the waterline (see Fig. 2). It is composed of three main sections: a surface tower fabricated from stainless steel, a central buoyancy section, and a subsurface spar section. The tower provides mounting locations for solar panels, ARGOS transmitters, radar reflector, and a navigation beacon. An instrument case, housing the electronics and data acquisition system, is mounted at the base of the tower. The buoyancy is provided by a Seaward International toroid float with 2950 kg total buoyancy. The center well of the toroid is filled with a battery pack connected to both the instrument case and the solar panels. The 7-m subsurface spar, fabricated of three sections of 203-mm-o.d. aluminum tubing, is the mounting point for the sound-speed modules and other sensors. A counterweight at the bottom of the spar provides additional stability. The three sections are designed to provide ease of assembly prior to deployment and simplify shipping of the buoy when disassembled.

Full scale drifter tests of the buoy system and its various components were carried out off Pt. Conception, California in an attempt to match the severe conditions expected in an Atlantic winter. Video studies of the buoy motion in 3.6-m seas revealed the buoy followed the longer waves (periods of 8 s and greater) with heave around ± 10 cm. In these tests, pitch and roll were limited to approximately 15°.

The travel-time technique employed on the sound-speed buoy is similar in principle to the design used by Lamarre and Melville.[25,26] The system was required to operate autonomously for a period of a few months, dictating some important aspects of the design: the resolution of the system is in the range of 2–5 m/s, measurements can be made simultaneously at seven discrete depths at a rate of 2 Hz, and the required computing power is relatively modest. Furthermore, the system is robust, a requirement for extended operation in the harsh conditions of the near-surface wave-breaking zone in the North Atlantic during the winter.

Each sound-speed module is composed of a rigid A-shaped frame with a transmit and receive transducer at the base of the ''A.'' Extensive laboratory acoustic testing of the A-frame and spar connection was conducted to avoid structural interference with the acoustic signals. An I.T.C. model no. 1032 transmit transducer and an I.T.C. model no. 1042 hydrophone were selected because of their resonant frequencies and response levels. The two modules nearest the surface have a 0.5-m pathlength to prevent multipath problems resulting from surface reflections. The deeper modules have a nominal pathlength of 1 m. Seven modules were spaced vertically at the following depths below the water line: 0.69, 1.02, 1.58, 1.96, 2.96, 4.44, and 6.94 m. The uneven vertical spacing of the modules provides a higher vertical resolution of the sound-speed field near the water surface where the largest gradients in sound speed are expected.

The sound speed is calculated through the direct measurement of the travel time of an acoustic pulse across the pathlength between the transmit and receive transducer.[25] The acoustic pulse is approximately three periods of a discrete frequency. The amplitude of the pulse is windowed to prevent impulsive loading of the transducer. The pulse is sent with a center frequency of 3.33, 5, or 10 kHz. The signal received at the hydrophone is transmitted via cable to the instrument case where it is digitized and processed.

Data acquisition was performed with a DAQBOOK (IOTECH Inc., Cleveland Ohio) parallel interface data acquisition system that was connected to a Compaq 486/33 notebook computer. The DAQBOOK obtains 128 samples of both the electrical signal sent to the transmit transducer and the received acoustic signal at a rate of 50 kHz per channel. The pulse travel times are calculated via cross-correlation techniques using fast Fourier transform (FFT) methods.[28] The travel times are stored on the computer's hard disk along with the received signal amplitudes.

II. EXPERIMENT DESCRIPTION

The Acoustic Surface Reverberation Experiment (ASREX) was a multi-investigator experiment designed to examine the effects of rough weather on sea-surface scattering and reverberation at frequencies below 1000 Hz. Several

instruments were deployed at both surface and subsurface locations housing a variety of acoustic and meteorological instruments. Participants in the experiment included scientists from the Institute of Ocean Science, Canada (IOS), Scripps Institution of Oceanography (SIO), University of Miami, and Woods Hole Oceanographic Institution (WHOI). The acoustic instrumentation consisted of the University of Miami's low-frequency source and vertical line array, upward looking and sidescan sonars deployed by the IOS group, and the Scripps sound-speed buoy. Environmental conditions were monitored by two surface buoys deployed by WHOI. The first was a Seatex Wavescan buoy which measures the surface gravity wave field through the interpretation of the buoy's pitch and roll data. It is able to resolve the directional wave field over wave periods ranging from 2 to 30 s. (Seatex, Trondheim, Norway).[29] The second buoy was a 3-m discus buoy (similar to those used by the National Data Buoy Center) which contained a full suite of meteorological sensors including wind speed and direction, sea-surface temperature, and barometric pressure. A detailed description of the experiment, including equipment summaries, deployment procedures, sampling strategies, and personnel involved can be found in Galbraith *et al.*[30]

The deep water mooring for the SIO buoy was designed and built by the Woods Hole Rigging Shop for the ASREX experiment (see Fig. 2). The mooring was designed to keep the buoy secure throughout the winter yet minimize any effects the mooring may have on the buoy's ability to follow the surface waves. The instrumented buoy was moored to a subsurface float at 500-m depth through an inverse catenary which decoupled the surface buoy from the mooring during periods of high currents. A scope (the ratio of horizontal distance to depth) of 2.62 was used between the surface buoy and the subsurface float at 500 m. The mooring was connected to the buoy through the use of a guard surface float that further decoupled the mooring from the buoy. The connection between the guard float and the buoy was composed of 30-m 3/8-in. chain covered by large diameter Tygon tubing. Small floats were connected to the surface tether at intervals of approximately 0.75 m to provide buoyancy. A similar design has been used in mooring the Seatex Wavescan buoy in deep water, successfully allowing the buoy to obtain unbiased wave measurements.

The moorings were set in the Atlantic at roughly 34 °N, 70 °W in mid-December 1993 using the R/V KNORR operated by WHOI. Nominal water depth was 4500 m. The mooring site was chosen for the flat bathymetry which allowed accurate placement of the subsurface instruments in the water column, and for the rough weather conditions typical of an Atlantic winter. After deployment of the sound-speed buoy, observations of the buoy motion under rough conditions became available as the winds quickly rose to 50 kn and peak wave heights were estimated to reach 12 m. Video of the buoy motion again revealed a maximum pitch/roll of approximately 15° and negligible heave. After a 3 month deployment, the moorings were recovered using the R/V EDWIN LINK operated by Harbor Branch Oceanographic Institution under more favorable seas.

For the 3 month deployment, the sound-speed buoy operated on a 44% duty cycle to extend the data storage capacity and battery power. The buoy operated with a schedule of 40 min on/50 min off with a travel-time measurement sample rate of 2 Hz per module. The center frequency of the acoustic pulse used for the measurement was changed at each new 40-min sampling period. Hence a 40-min time series of 3.33-kHz sound-speed data would be available every 4.5 h over the duration of the deployment. Sound-speed averages, battery voltages, and buoy position were telemetered to shore via the ARGOS system on a daily basis.

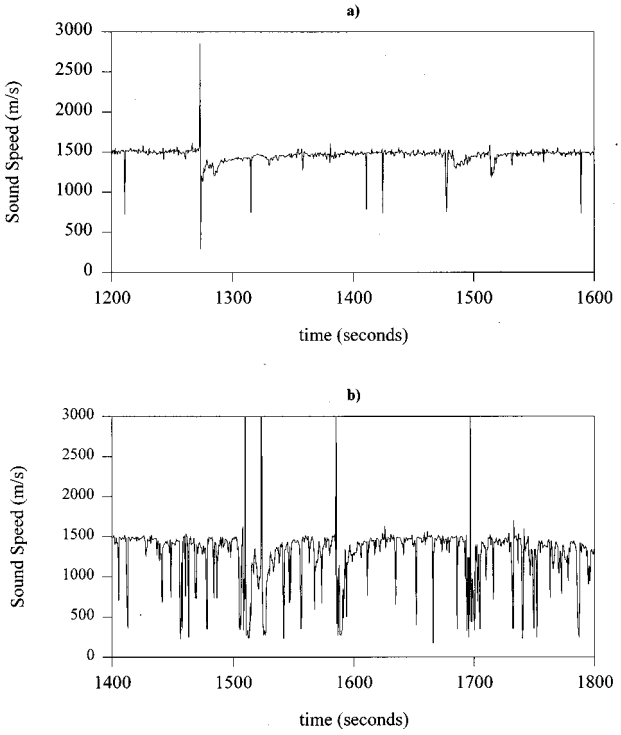


FIG. 3. Example time series of 3.33-kHz sound speed, for (a) moderate sea state ($H_{1/3}=2.3$ m, $U_{10}=11.2$ m/s) and (b) extremely rough conditions 13.5 h later ($H_{1/3}=4.2$ m, $U_{10}=12.0$ m/s). Bad data points include those that exist at values much greater than the bubble-free sound speed as well as the single points that exist near 700 m/s.

III. DATA

A. Data available

Upon recovery of the buoy, it was discovered that only 43 days of data were recorded, somewhat less than the desired 3 months. Close inspection of the system voltage over the duration of the deployment showed a sudden drop from which the buoy never recovered on 21 January 1994. Postrecovery inspection of the buoy showed evidence of abrasion in one of the submerged cables, resulting in exposure of the conductors to seawater. Other nearby instruments also experienced similar cable degradation to varying degrees. This was attributed to fish bite. With the conductors grounded to seawater, the buoy's batteries drained below acceptable system operating levels.

Directional wave data from the Seatex buoy and meteorological observations from the 3-m discuss were successfully obtained by WHOI over the interval of sound-speed

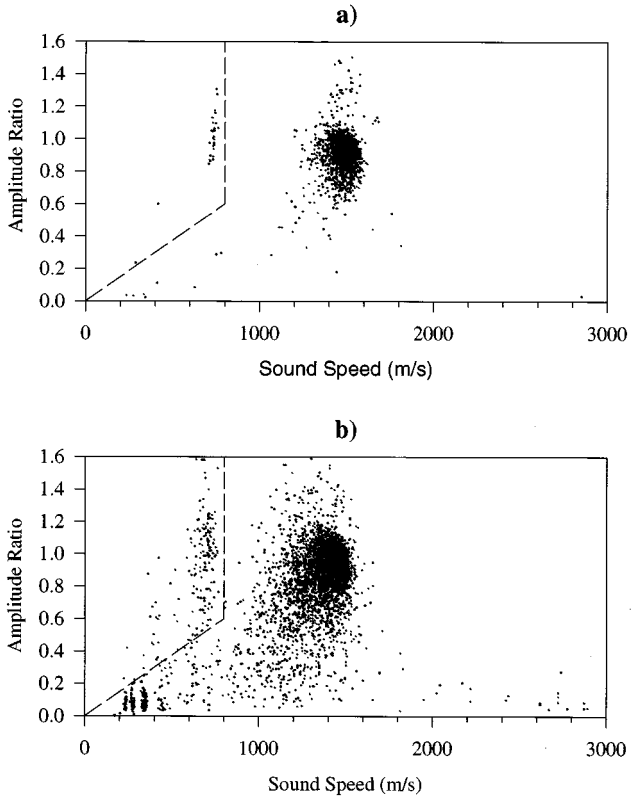


FIG. 4. Scatter plots of sound speed versus received signal amplitude for both the previously shown time series during (a) a moderate sea state and (b) rougher conditions. The dashed line is the criterion used to remove bad data points.

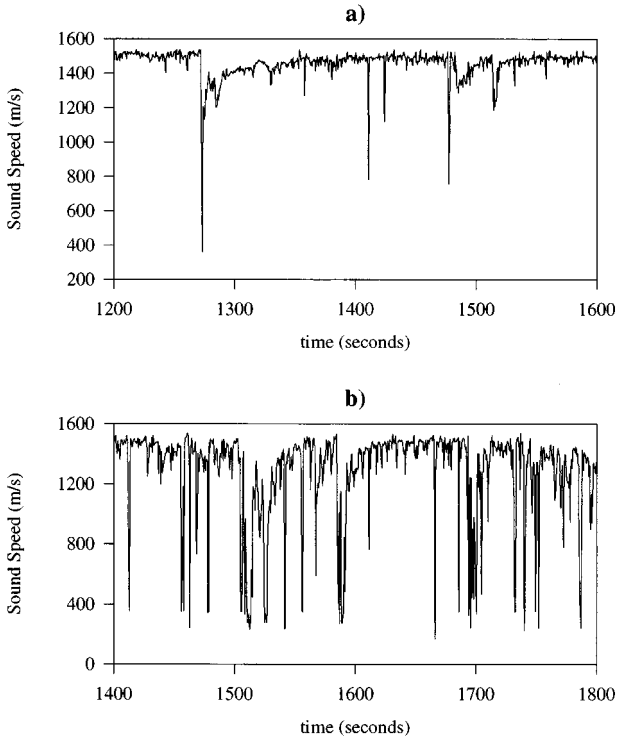


FIG. 5. The sound-speed time series of Fig. 3 after the bad data points are removed and replaced with the average of the neighboring points. Again, (a) is the moderate sea state and (b) is the rougher conditions.

measurements, providing the necessary environmental measurements for comparison with our data set. The meteorological measurements show that for the 43 days of operation, seven intense storms lasting 2 to 3 days passed over the array of moorings. All storms had winds exceeding 15 m/s accompanied by significant wave heights of several meters or more.

B. Preprocessing of acoustic data

In order to calculate the sound speed from the travel-time data, accurate measurements of the acoustic pathlength are needed for each sound-speed module. This pathlength includes any differences between the physical and acoustic centers of the transducers, and any changes in the separation of the transducers due to unforeseen incidents (e.g., fish hits). A convenient method of measuring the pathlength is to use travel-time data obtained during very calm periods, when few bubbles exist in the water column. Since the bubble-free sound speed can be accurately calculated using temperature and salinity information,[31] an *in-situ* acoustic pathlength may be computed during various times throughout the deployment. The acoustic pathlengths were computed using four calm periods in the 43 day record and found not to vary by more than $O(0.001)$ m for any of the seven modules. Resolving the standard deviations of the pathlength measurements as a sound speed, the sound-speed module's error due to variations in pathlength would be approximately ± 3 m/s over the course of the measurements. These uncertainties are similar in magnitude to the resolution of the system due to sampling rates of the acoustic signals.

As with any new instrument, the quality of the sound-speed data required assessment. It was expected that during extreme weather, ambient noise and signal attenuation may cause the signal-to-noise ratio to drop below acceptable levels for accurate travel-time measurement. To what extent this would occur was unknown because of the novelty of the measurement under these extreme conditions. After examining the data for the seven depths, it appeared that data from the two shallowest measurements contained some spurious points during periods of intense breaking. Figure 3 shows two time series of 3.33-kHz sound-speed data measured at 0.7 m during a storm event. Time series (a) is representative of the sound speed during a moderate sea state ($H_{1/3} = 2.3$ m, $U_{10} = 11.2$ m/s) while time series (b), obtained 13.5 h later, represents much rougher conditions ($H_{1/3} = 4.2$ m, $U_{10} = 12.0$ m/s). Two features present in both time series, thought to be bad data points, are the isolated points at levels much greater than the bubble-free sound speed ($\simeq 1500$ m/s) as well as single data points that seem to occur near 700 m/s.

Some insight into recognizing bad data points of the time series is possible if scatter plots are made of the sound speed against the ratio of the amplitude of the received to transmitted signal. Figure 4 shows the corresponding scatter plots for the time series in Fig. 3. Three distinct clusters appear in the two plots: good data consisting of the majority of the points and two groupings of questionable data resulting from different mechanisms. The first grouping of bad

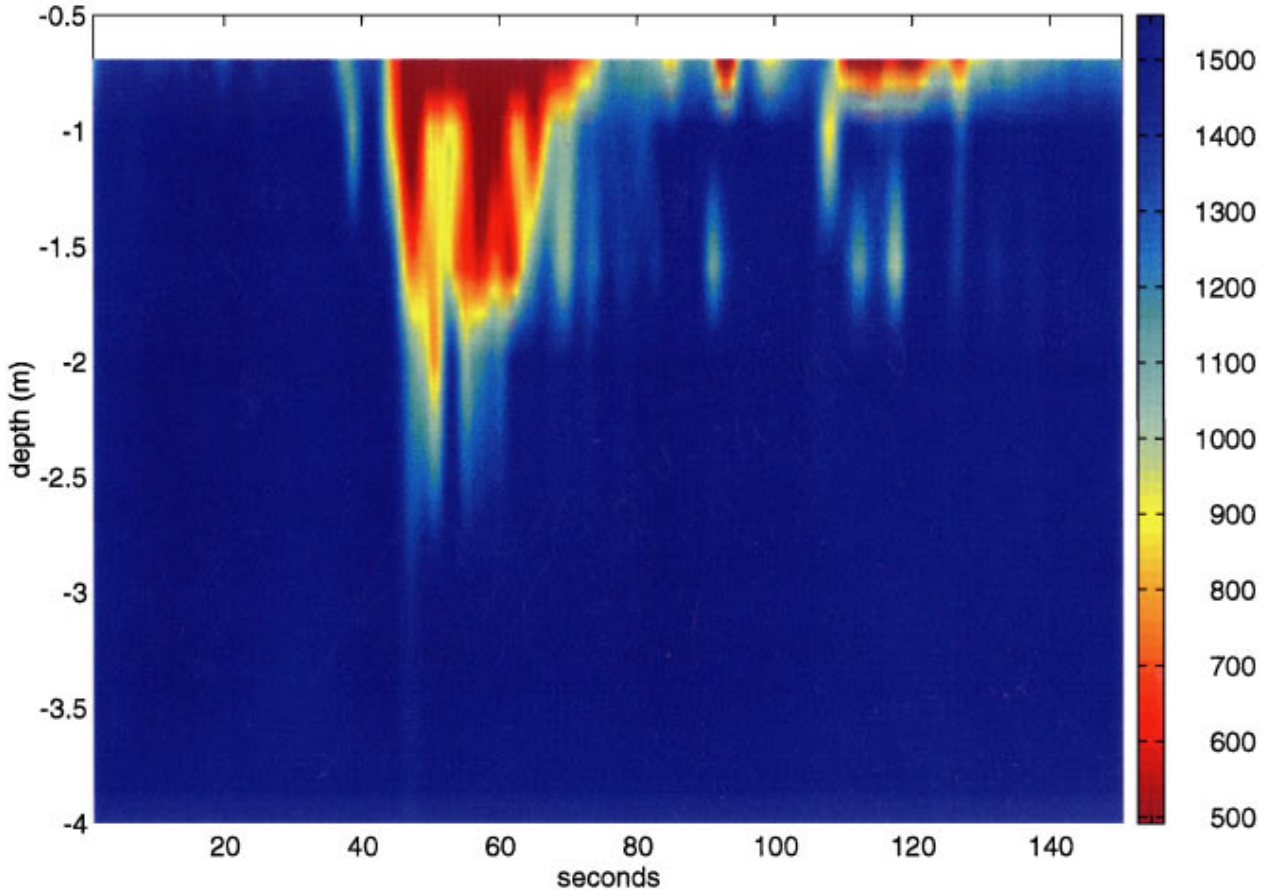


FIG. 6. Color contour plot of the response of the sound-speed field to a large single event. A region of very low sound speed exists near the surface for $O(30)$ s while at depth, the reductions are smaller and last for shorter periods of time. The high-frequency oscillation visible in the figure is attributed to orbital motions of the waves advecting the bubble cloud past the sensors.

data is composed of the points of very large sound speed but with small amplitude. These points are attributed to a low signal-to-noise ratio (SNR), i.e., the signal level is too low for the cross correlation to effectively find the arrival time of the transmitted acoustic wave. If the transmitted acoustic pulse is masked by ambient noise or attenuation, low-level electrical cross talk between the transmit and receive transducer cables may be cross correlated, resulting in very short travel times (large sound speeds).

The second group of the bad data points are those that appear shifted ~ 700 m/s from the central cluster of good points. These points occur because the peak of the cross correlation is shifted in time by one period in the three period acoustic pulse. The resulting calculated sound speed is then

$$C = \frac{L}{T + T_\lambda}, \tag{1}$$

where T_λ is the period of one 3.33-kHz wavelength. Using scatter plots of sound speed and amplitude over the 43 days of data, an empirical criterion for bad data points was arrived and that effectively separates the bad data results from the erroneous phase shifting. The dashed lines shown in Fig. 4 show boundaries by which good data is separated from bad data. The criteria were empirically selected after analyzing similar scatter plots over the 43 day record. Careful examination of the bad data points reveals that they usually are single-point anomalies in the 2-Hz data (i.e., they are discontinuous with their neighboring points), providing confidence in the criterion used to identify bad data. Once the bad data points are identified, they are replaced with the average value of their neighbors, providing a gap-free data set (Fig. 5).

A final feature present in the scatter plots of Fig. 4 that deserves mention are the groupings that begin to appear near 360 m/s. These data points are found to exist during large events of sound-speed reduction, lasting for several seconds before the measured speeds begin to rise, presumably reflecting the presence of high void fractions and subsequent degassing. The groupings arise as a consequence of the acoustic pulse delayed beyond the width of the analog-to-digital sample window, resulting in the received pulse being clipped in time. When this occurs, the time delay which results from the cross-correlation peak may result in erroneous values. Grouping arises from the cross-correlation peak shifting by 2π, dependent on the integral number of wavelengths the pulse is delayed outside the window.

In light of previous laboratory work, it is not surprising that the near-surface measurements are reduced to levels below 360 m/s, which corresponds to void fractions in excess

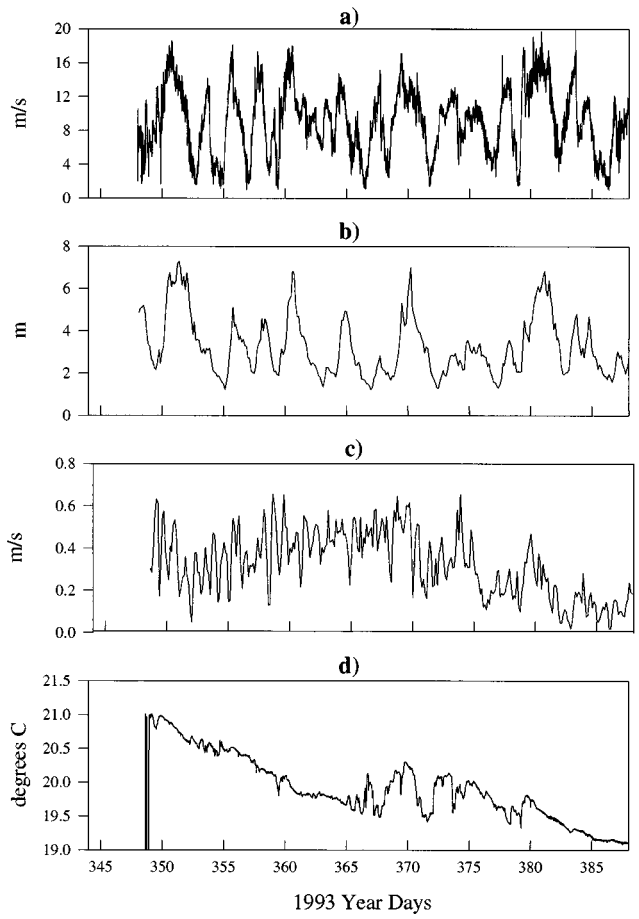


FIG. 7. Time series of the environmental conditions while the sound-speed buoy was operational: (a) Wind speed corrected to 10-m height. (b) Significant wave height. (c) Surface currents measured at 5-m depth. (d) Sea surface temperature measured at 1-m depth.

of 10^{-3}. Lamarre and Melville[24] show that void fractions can exceed 50% for short periods directly beneath a laboratory generated wave. Similarly, it is likely that void fractions would approach or exceed 50% for short periods of time in the immediate vicinity of large breakers considering the extreme weather encountered during ASREX. To resolve the effects of the instrumentation's inability to measure very low sound speeds on the calculated mean sound speeds, conservative error bounds are set based on Wood's curve and the instrument's sample window. An upper bound is set if the ambiguous data points below 360 m/s are conservatively set equal to the 360-m/s level that is confidently resolved. The lower bound is set based on assuming that the lowest values the ambiguous points would approach is the lower limit of Wood's curve: at high void fractions of approximately 50% a sound-speed minimum exists at 20 m/s. The minimum reflects the influence of the mixture's density on the sound speed at high void fractions. At very high void fractions, the curve approaches the sound speed of air. For the rough sea state conditions shown in Figs. 3(b), 4(b), and 5(b), the resulting range of mean sound speeds bounded by these limits is 1284–1299 m/s, a 1% error in the mean sound speed and a 6% uncertainty in the mean sound-speed reduction.

As expected, the uncertainty of the mean sound speed is related to the quantity of data that exceeds the lower limit of the sound velocimeter over the course of a measurement cycle, which in turn, is dependent on the sea state and measurement depth. Typically at the 0.7-m measurement depth during rough weather, uncertainty in the mean reduction is 10–20 m/s or $O(5)$%. The next measurement depth at 1 m typically contains $O(1)$% uncertainty while the deeper measurements always contained much less than 1% bad data, resulting in negligible uncertainty of the mean reduction. For the remainder of this paper, the reported mean sound speeds at the shallowest depths are the larger (more conservative) of the two values (the smaller mean reduction).

While the average data reflect the results of rough weather on the sound-speed field, they do little to show how the field responds to individual breaking events. Figure 6 shows a color contour plot of the sound-speed field in the presence of a large bubble cloud that resulted from a nearby breaking wave. Time in seconds is shown on the horizontal axis, depth in meters is represented on the vertical axis, and sound speed is represented by color. Pixels between the seven measurement depths have been linearly interpolated to arrive at a smoothed figure. For this particular event, the sound speed is drastically reduced at the surface to $O(500)$ m/s for 30–40 s. A few meters deeper, reductions of $O(100)$ m/s appear but tend to last for shorter time periods. It is short-lived events like the one shown (also visible in later figures) that are primarily responsible for injecting air into the water column and lowering the mean sound speeds.

IV. PROCESSED DATA: 40-MIN AVERAGE SOUND SPEEDS

A. 3.33-kHz data summary

Successful operation of the sound speed buoy for 43 days provides data for analysis over a wide range of environmental conditions. Figure 7 presents time series of wind speed, significant wave height, surface currents, and sea surface temperature measured at the directional wave buoy and the 3-m meteorological buoy. Figure 8 shows the 40 min average sound speed for 3.33 kHz at the seven different depths along with significant wave height. The sea surface temperature, measured at 1 m is also plotted with the deepest sound-speed measurement.

It is apparent that the layer of water near the ocean's surface exhibits large gradients in sound speeds. Near the surface, average sound speeds routinely drop to approximately 1200 m/s, in contrast to the 7-m depth, where the average sound speed appears to follow secular trends in the temperature field. The average data also depicts the variability that exists in the near-surface sound-speed field from one storm to another. This is exemplified by the suppressed response of the sound-speed field for the storm that occurs year day 380, 1993 in comparison with the storm that occurs just ten days earlier on day 370. Note that both storms have similar peak wind speeds and significant wave heights.

B. Average sound-speed dependence on environmental parameters

If the presence of air in the near-surface waters is due to wave breaking, we need to consider which environmental

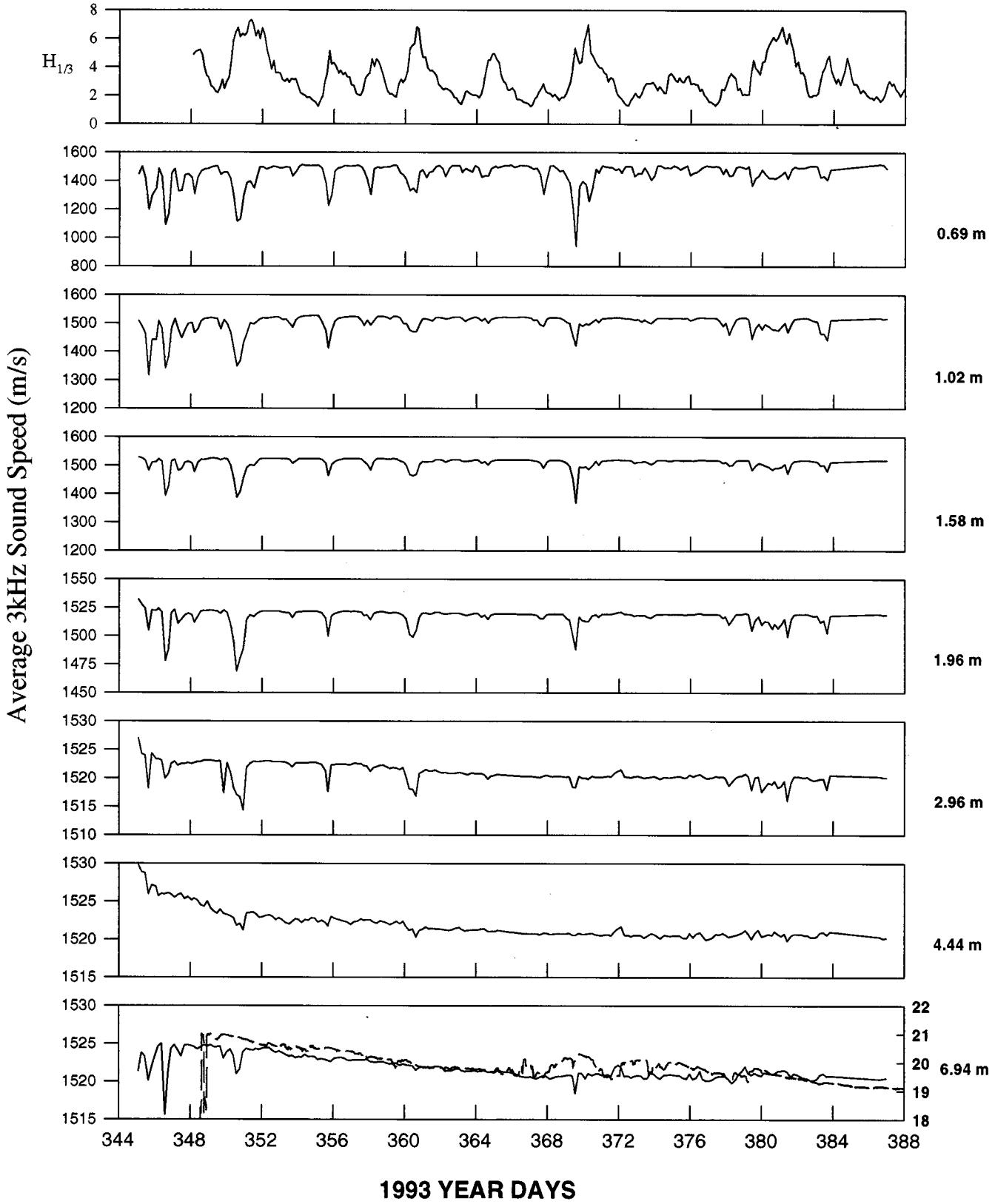


FIG. 8. Time series of the mean 3.33-kHz sound speeds measured at the seven measurement depths for the duration of the deployment. Time series of significant wave height are provided to note the locations of storms. Sea surface temperature is plotted with the deepest measurement depth to show similar trends. Note the varying y-axis scale for the different measurement depths.

conditions control deep-water wave breaking. Historically, the wind speed and significant wave height have been used for obtaining correlations with bubble-driven acoustic phenomena such as backscatter and ambient noise, with limited success. In a comparison of ambient noise with surface wave variables, Felizardo and Melville[32,33] point out that the presence of swell can significantly alter the measured significant wave height while only contributing second-order effects to

the wave breaking, or the acoustically active portion of the wave field. Their work suggests that other parameters exist that better represent the intensity of wave breaking in the surface wave field. These include the wave slope and dissipation of energy of the surface wave field. The average sound-speed reductions obtained in ASREX are compared with both the classical environmental parameters of wind speed and significant wave height as well as with parameters which better describe the dynamics of the sea surface.

Figure 9(a) and (b) compares the average sound-speed reductions with the average wind speed, corrected to 10-m height, and the significant wave height. Averages are calculated from 40 min time series of data obtained from the sound-speed modules located at depths of 0.7 and 1.6 m. The two depths are selected because the shallowest measurement is in the most active region of the water column and the deeper depth is representative of the mid-depth signal of the bubbly layer. As expected, correlations exist between the average sound speeds and the wind and wave field, albeit with significant scatter.

At windspeeds less than 8 m/s, the near-surface measurement at 0.7 m appears to be independent of the windspeed while the deeper measurement appears to be sensitive across a wider range of windspeeds. The figures reflect the variability in the shallow entrainment of air, particularly at lower wind speeds. Figure 9(b) presents a comparison of the mean reductions with the significant wave height. While a coupling of the wave field to the wave breaking process is suggested by the trends, the scatter suggests a need for a better descriptor based on the wave field. In reality, the rate and intensity of surface wave breaking, which controls the quantity of air injected, will be governed by a complex combination of environmental conditions that contribute to the sea state.

Laboratory experiments by Rapp and Melville[34] and Loewen and Melville[35] have suggested that the slope of the wave field can be used as an indicator of the strength of breaking. Furthermore, their results suggest that the steepness of the field may reflect the levels at which other related dynamic processes are occurring. This includes the associated momentum flux, mixing, and dissipation of the surface waves, as well as acoustic phenomena related to wave breaking such as ambient noise generation. Felizardo and Melville's[32] field measurements support the laboratory experiments by showing strong correlations of ambient noise in various frequency bands with the steepness of the wave field. Using a four wire wave gauge array, Felizardo and Melville[32] computed time series of the slope of the wave field from the finite difference of wave elevation between wires in the array. Comparisons of rms slope with measured ambient noise levels show a strong correlation over the 0–12 m/s wind speeds encountered over the course of that experiment.

In light of the previous results, it is useful to derive an estimate of the steepness of the wave field during the ASREX experiment. Unfortunately, the wave measurements were obtained using a pitch–roll buoy, providing no direct measurement of the wave slopes. In order to arrive at a first-order estimate of the steepness of the wave field, information from the surface displacement power spectra is used. The power spectra of the surface elevation $\Phi(\omega)$ can be converted to a slope spectrum $B(\omega)$ through the linear dispersion relationship. Assuming the dispersion relation $\omega^2 = gk$, the corresponding slope spectrum is

$$B(\omega) = \frac{\Phi(\omega)\omega^4}{g^2}. \tag{2}$$

The slope spectrum is then integrated over a range of frequencies to arrive a parameter,

$$S_p = \int_{\omega 1}^{\omega 2} B(\omega)\, d\omega, \tag{3}$$

that is a measure of the steepness of the wave field. The low-frequency limit is set to correspond to 15-s waves, ensuring that the majority of the wind–wave spectrum is incorporated into the integral. The high-frequency limit of the integral is set to wave periods of 3 s, corresponding to the upper frequency limit at which the pitch–roll buoy appeared to accurately resolve the surface waves. Figure 9(c) shows the comparison of the slope parameter with the average sound-speed reduction, again for the 0.7- and 1.6-m depth measurement. The figure highlights the better correlations this parameter provides with the sound-speed reductions when compared to the significant wave height comparisons of Fig. 9(b). While the value of the slope parameter S_p does vary with the limits of integration, the resulting trends are insensitive to the integration limits if the frequency range covers the wind/wave spectrum.

The final parameter used for comparison with the sound-speed data is an estimate of the energy dissipated by the surface wave field. In the work of Felizardo and Melville,[32] estimates of the energy dissipation of the surface field[36–38] are found to correlate well with the portion of the ambient noise spectrum that results from wave-injected bubbles. If the energy dissipation of the surface wave field is due to wave breaking, it is expected to correlate with sound-speed reductions due to air injection. Earlier work by Melville *et al.*[39] and Loewen and Melville[35] in the laboratory and Kennedy[40] in the field produced results showing that the sound radiated by breaking scales with the energy dissipated by wave breaking.

To investigate the dissipation term, an existing model is used which is based on bulk parameters. A quasi-equilibrium model proposed by Phillips[36] assumes a balance between the nonlinear wave interactions, wind input, and wave breaking dissipation over an equilibrium range of wave numbers. Using previous models of the wind input and nonlinear transfers, he showed that the spectral rate of energy loss from the waves (within the equilibrium spectral range) was given by $\epsilon(k) = \gamma\beta^3(\cos\theta)^{3p} u_{*_a}^3 k^{-2}$, where γ, β, and p are numerical constants, θ is the included angle between the wave-number vector k and the wind vector, and u_{*_a} is the friction velocity of the air flow over the water surface. Integrating over a range of wave numbers with a lower limit at the peak of the wave spectrum and an upper limit (high-frequency waves)

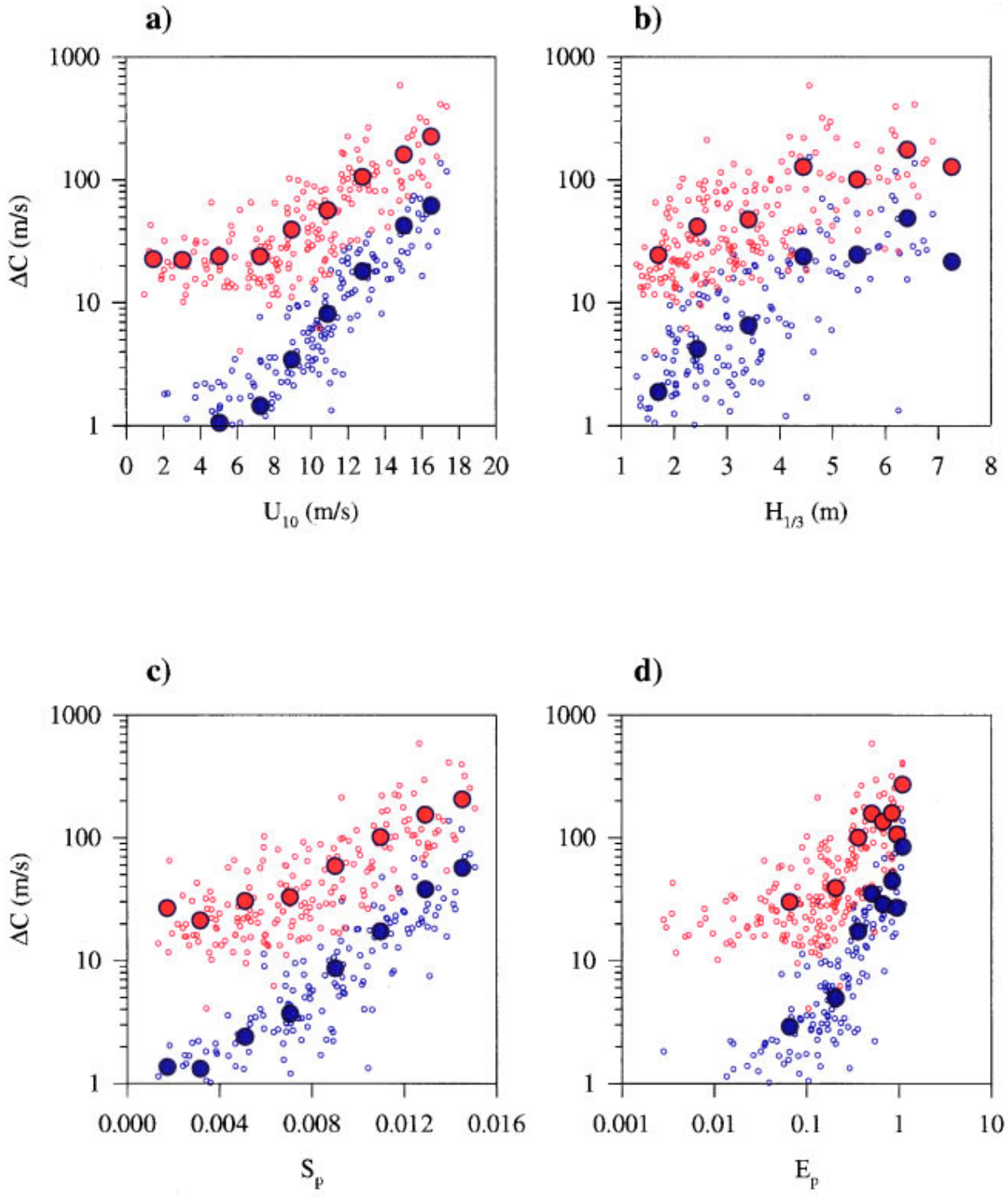


FIG. 9. Comparisons of the mean sound-speed reductions with (a) wind speed corrected to 10 m above the sea surface, (b) significant wave height, (c) computed slope parameter, (d) Phillip's (Ref. 36) bulk dissipation estimate. Data from the 0.7- and 1.6-m measurement depths are represented by the colors red and blue, respectively. Solid circles represent averages of the data grouped with the following binwidths: (a) 2 m/s, (b) 1 m, (c) 0.002, (d) 0.01.

which is determined by suppression of waves due to the presence of wind drift, a bulk estimate of the dissipation of the wave field arises and is given by

$$D = 2\gamma\beta_3 I(3p)\rho_w u_*^3 \ln\left[r\left(\frac{c_p}{u_*}\right)^2\right]. \quad (4)$$

Felizardo and Melville[31] showed that the range of the numerical constants $\gamma\beta^3 I(3p)$ is $3.7-8.0\times10^{-4}$; a factor of approximately 2. The advantage of comparing this equation with measured values of entrained air is that the dissipation estimate is based not only on the wind speed (friction velocity u_*) but the wave age c_p/u_*, which incorporates information about the surface wave spectrum (c_p is the speed of the waves at the peak of the wave spectrum). Hence the bulk estimate of the energy loss of the wave field that is important to wind–wave modeling[41] provides a parameter that is de-

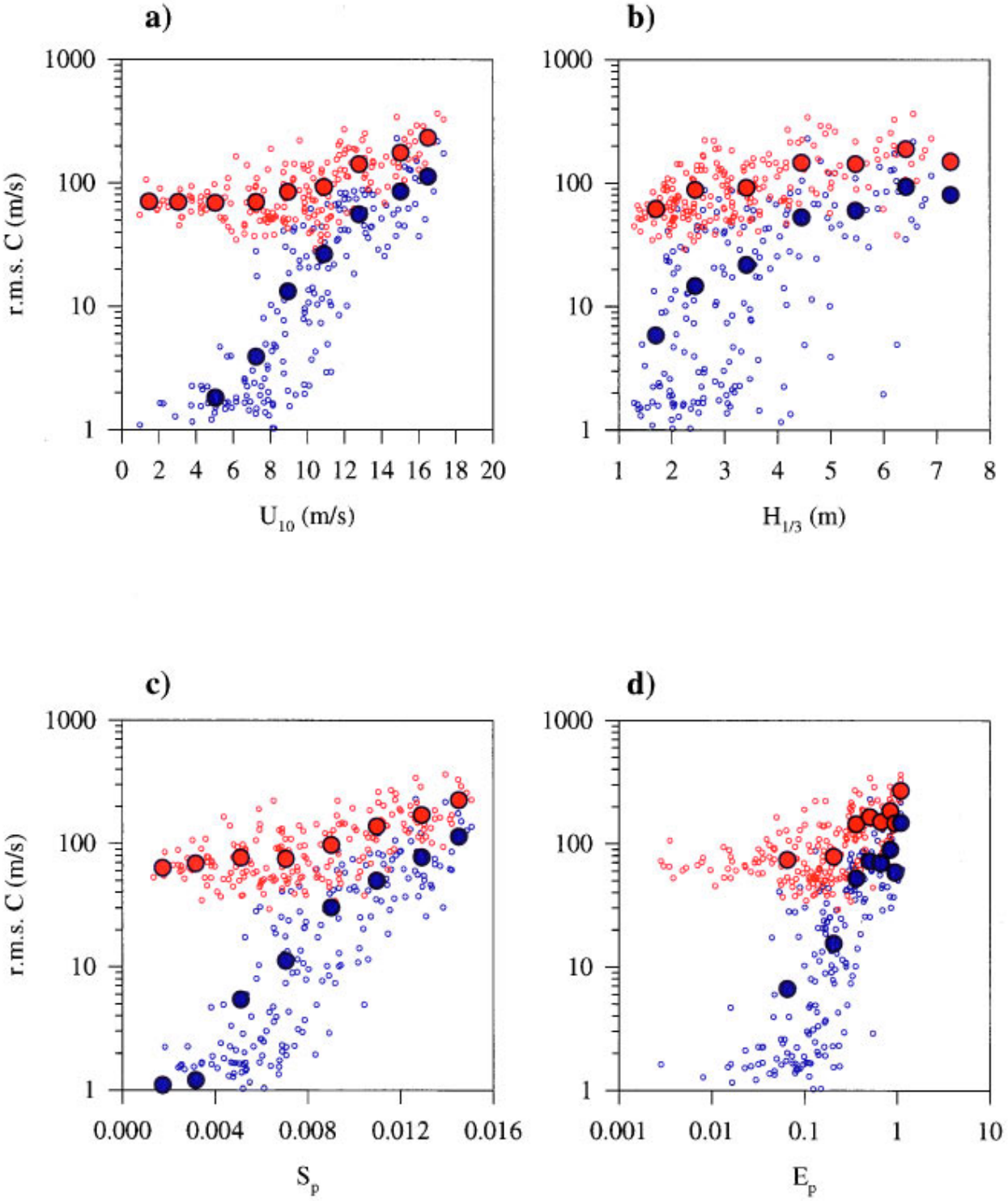


FIG. 10. Comparisons of the rms sound speed with (a) wind speed corrected to 10 m above the sea surface, (b) significant wave height, (c) computed slope parameter, (d) Phillip's (Ref. 36) bulk dissipation estimate. Data from the 0.7- and 1.6-m measurement depths are represented by the colors red and blue, respectively. Solid circles represent averages of the data grouped with the following binwidths: (a) 2 m/s (b) 1 m, (c) 0.002, (d) 0.01.

pendent on both the wind and wave conditions which can be compared with measurements of bubble-driven acoustic phenomena.

Dissipation estimates based on the wind and wave conditions during the ASREX experiment have been computed and compared with the 40-min average sound-speed deviations at both 0.7- and 1.6-m measurement depth, as is shown in Fig. 9(d). It becomes evident that the average sound speed correlates with the dissipation estimate. While the scatter in the data suggests room for improvement in the dissipation estimate, it provides a rational means by which environmental variables are combined to arrive at a description of the intensity of wave breaking at the ocean's surface. The scatter is not surprising in light of the fact that the dissipation term

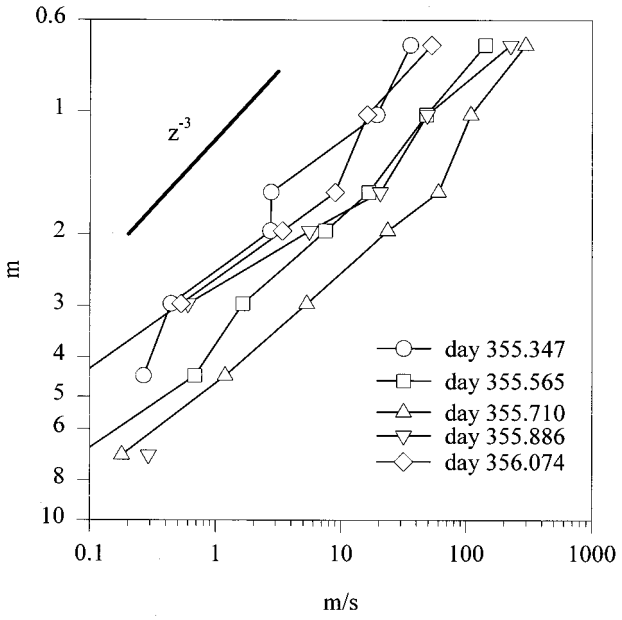


FIG. 11. Depth profiles of the mean sound-speed reduction measured over the course of a storm that began 1993 year day 355. Note that all profiles vary approximately as z^{-3}.

is the least understood of the terms in the radiative transfer equation[36,41–43] and presently does not include the effects of swell and wave-current interactions. It is expected that improved formulations of the dissipation function would improve the acoustic comparisons. Nevertheless, the dissipation estimate provides a descriptor of the ocean's surface which has a sound physical basis for comparison with acoustic observations.

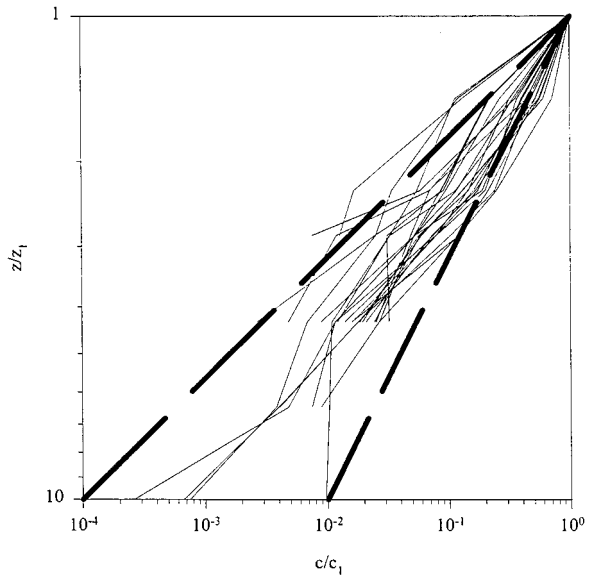


FIG. 12. Several example mean sound-speed reduction depth profiles for rough conditions measured over the course of the experiment. The profiles are normalized by the depth and mean sound speed of the shallowest measurement. Power law slopes of -2 and -4 are given by the heavy dashed lines.

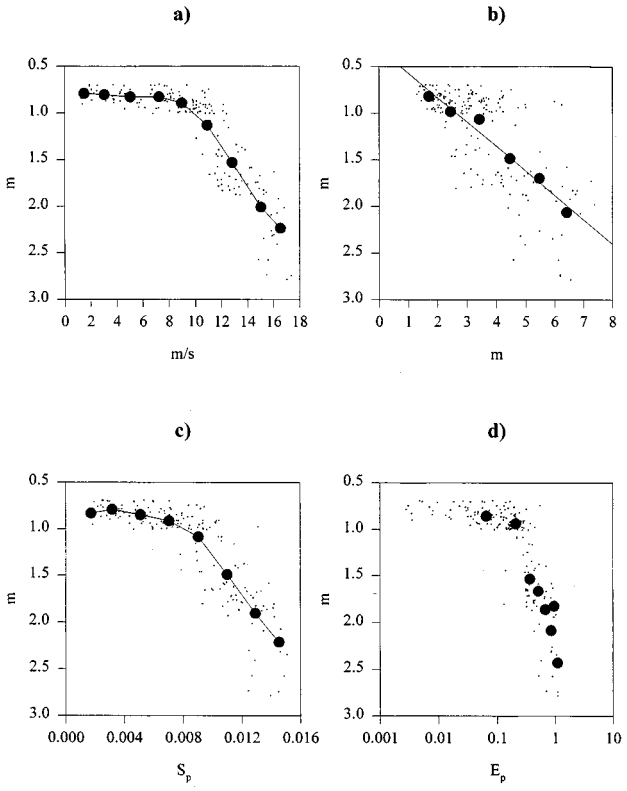


FIG. 13. The depth representing the level at which the mean sound speed is 99% of the bubble-free value ($\beta \cong 10^{-6}$) is compared with (a) wind speed corrected to 10 m above the sea surface, (b) significant wave height, (c) computed slope parameter, (d) Phillip's (Ref. 36) bulk dissipation estimate. The solid circles represent averages of the data grouped with the following binwidths: (a) 2 m/s, (b) 1 m, (c) 0.002, (d) 0.01. Note the linear relationship between the contour depth and the significant wave height (binned).

C. The rms sound-speed dependence on environmental parameters

The rms of the 3.33-kHz sound-speed data is analyzed in similar fashion to the mean data. Again the analysis is limited to the 0.7- and 1.6-m depth measurements. Figure 10(a) and (b) shows the rms sound speed at the two measurement depths compared with the windspeed and significant wave height. The trends that appear in the comparisons closely resemble those of the comparisons made with the mean sound-speed data (Fig. 9). Significant scatter again exists between the rms sound speed at the shallowest measurement with both the wind speed or wave height variable. Furthermore the scatter is reduced if the environmental variable is compared with the deeper measurement (note that ΔC is plotted on a logarithmic scale). The rms sound speed is also compared with the wave steepness parameter and the wave dissipation estimate that was discussed above [Fig. 10(c) and (d)]. The strong correlations again suggest a coupling of the surface wave-field dynamics to the sound-speed field and that other parameters exist that may better describe the intensity of wave breaking.

D. Depth dependence of average sound speed

In principle, the buoyant effects of bubbles ensure that the average sound-speed reductions will be at a maximum near the ocean's surface and monotonically decrease with

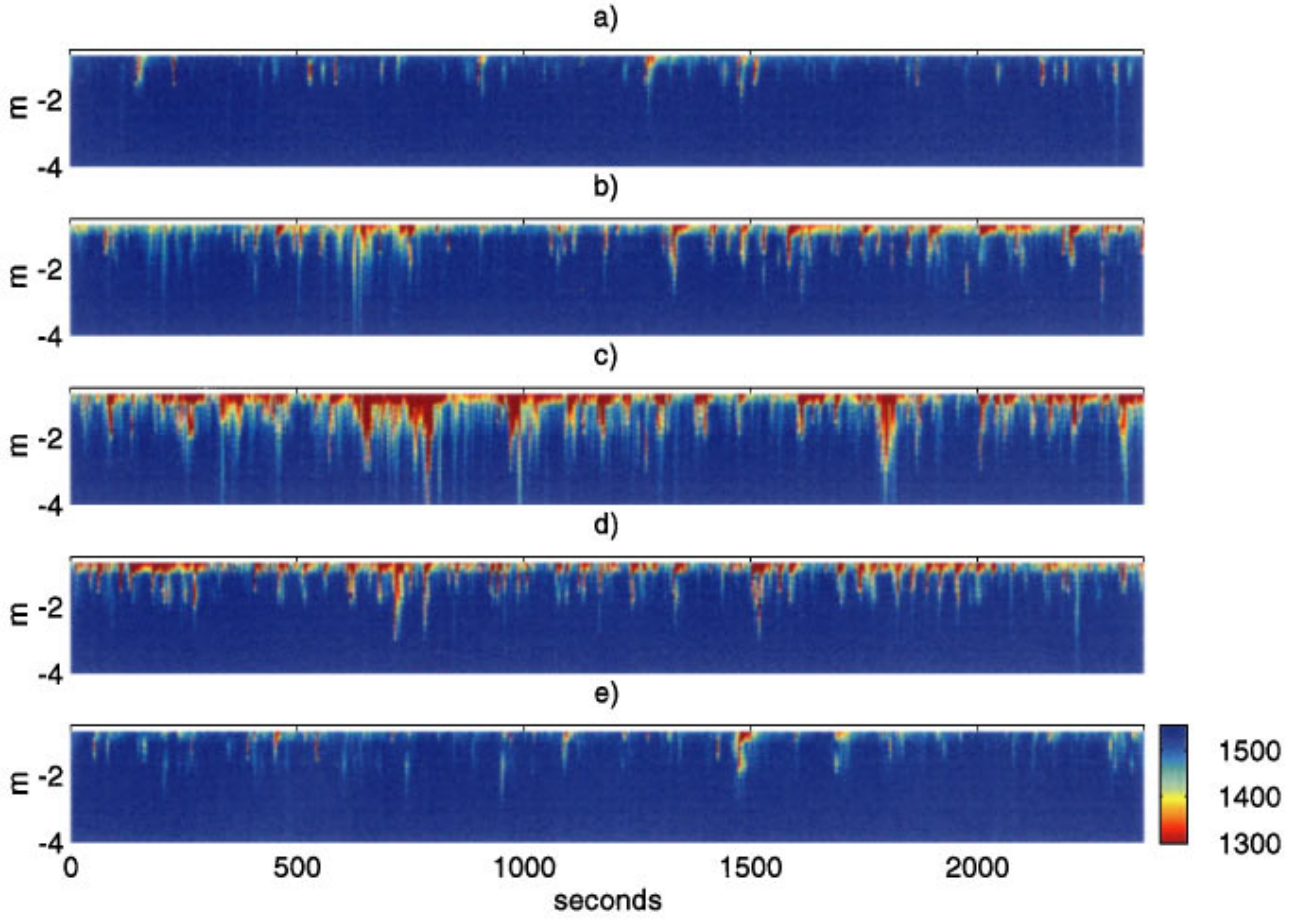


FIG. 14. Color contour plots of the 3.33 kHz sound-speed time series during a storm. Each figure is separated by 4.5 h and corresponds to the data shown in the depth profiles of Fig. 11. The 2-Hz sampled data is linearly interpolated with depth to arrive at a smoothed figure.

depth. In addition, the flux of bubbles into the surface waters increases as the incidence of wave breaking increases, further reducing the sound speed. This qualitative description of the near-surface sound-speed field has been observed in the earlier measurements of Farmer and Vagle[22] and Lamarre and Melville.[25] The large data set provided by the ASREX deployment also exhibits this trend of a decreasing sound-speed reduction depth profile, even during the roughest of conditions. Figure 11 presents the depth profiles of the mean sound-speed reductions measured over the course of a storm that began on year day 355, 1993. Each profile represents 40 min of data separated by 4.5 h. It is apparent that the average sound-speed profiles respond to the sea state, with the peak of the storm reflected in the profile with the largest sound-speed reductions.

The linear trend exhibited on the log-log scale suggests that a power-law relationship is the appropriate functional relationship for the depth dependence of the data. Furthermore, despite variations in the magnitude of the data over the course of the storm, the slope of the power-law relation appears almost constant with a value of approximately -3. This contrasts with the previous work of both Farmer and Vagle[22] and Lamarre and Melville[25] who used exponentials to describe the depth dependence of their sound-speed observations. A number of profiles spanning the 43-day record are normalized with respect to the shallowest measurement and presented in Fig. 12. The profiles shown, which are representative of the data obtained over the course of the experiment, again exhibit power-law relations with slopes in the range -2 to -4.

To explore the depth dependence of the layer of reduced sound speed, a contour representing the depth at which the mean sound speed is 99% of the bubble-free value is calculated ($z_{0.99C_0}$). It's value is computed by linearly interpolating between the mean sound-speed levels at each of the seven measurement depths. The depth of the layer is compared with the various environmental parameters used to described the mean sound-speed levels [Fig. 13(a)–(d)]. The resulting trends point towards a layer depth that increases with both the wind speed and significant wave height as well as with the estimates of wave steepness and wave energy dissipation. The data suggest a linear relationship between the depth of the reduced sound-speed layer and the significant wave height:

$$z_{0.99C_0} = 0.26H_{1/3} + 0.31 (r^2 = 0.97), \tag{5}$$

where r^2 is the correlation coefficient [regression performed after binning the data according to 1-m increments in wave height, see Fig. 13(b)]. Since the significant wave height is approximately four times the wave amplitude rms,[41] the re-

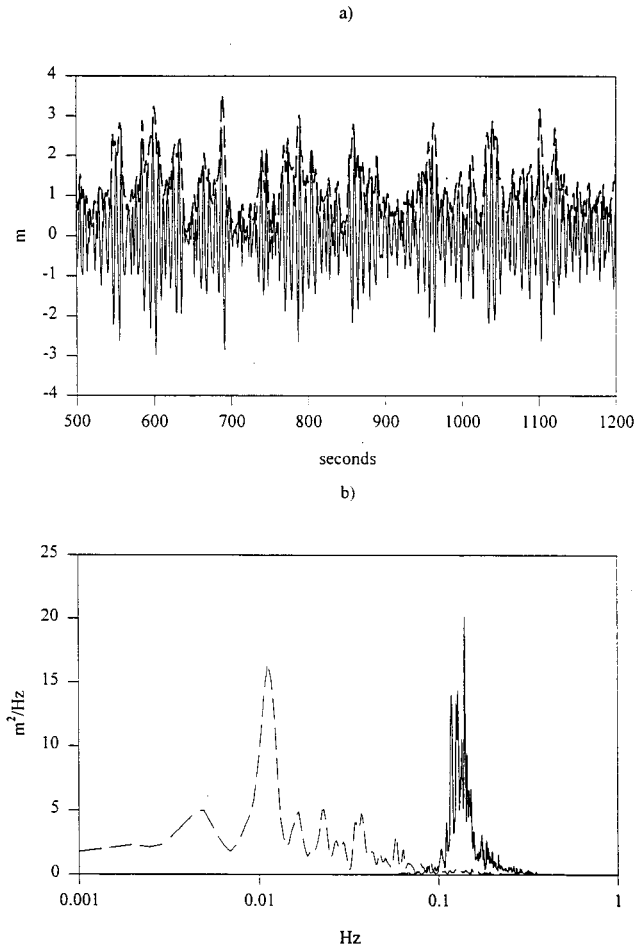


FIG. 15. (a) An example time series of the surface elevation (solid line) and the envelope function (dashed line) computed through the use of the Hilbert transform. Note the groupiness of the wave field, outlined by the wave envelope. (b) Computed power spectra of the surface wave heights (solid line) and the power spectra of the envelope function (dashed line). Distinct peaks occur at the peak of the wind wave spectrum at $O(0.1)$ Hz while the wave group frequency appears to have a peak at $O(0.01)$.

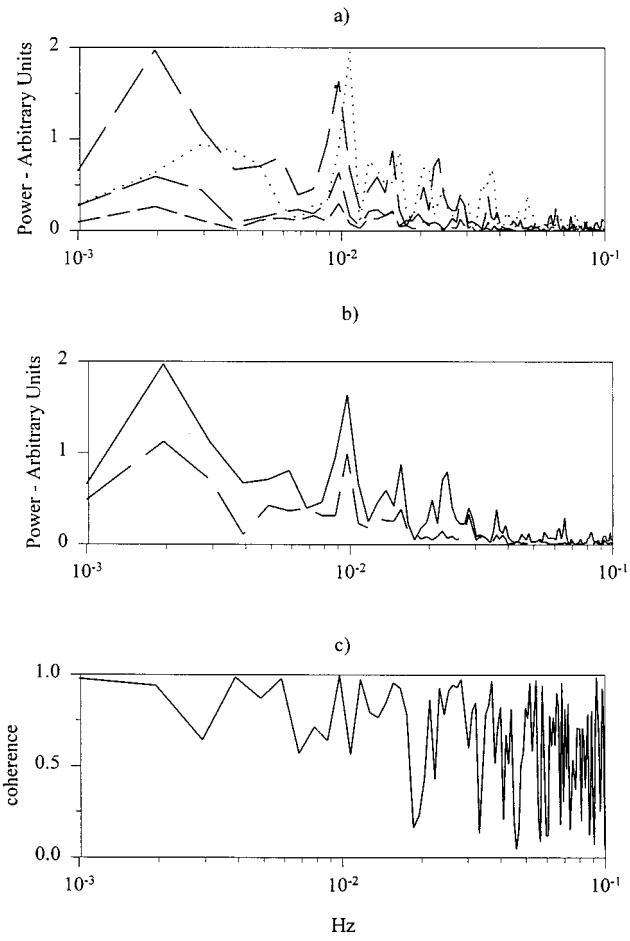


FIG. 16. (a) Spectra of the sound-speed field obtained from the shallowest three measurements and the spectra of the wave envelope (long dash—0.7 m, medium dash—1.0 m, short dash—1.6 m, dotted line—wave envelope). Good agreement between the spectra appears at $O(0.01)$ Hz, the peak of the wave envelope spectrum. Vertical axis linear. (b) Spectra of the sound speed measured at 0.7 m (solid line) and the contour representing the depth (dashed line) at which the sound speed is 90% of the bubble-free value ($\beta \cong 10^{-5}$). Vertical axis linear. (c) Computed coherence between the contour depth and and sound-speed measured at 0.7 m. High coherence exists between the two at the lower wave group frequencies.

sults suggest that the 99% contour depth has a characteristic length that scales with the wave amplitude rms.

V. FINE SCALE PROPERTIES OF SOUND-SPEED FIELD

In order to appreciate the nature of the sound-speed field at smaller time scales, it is helpful to analyze the raw time series that were obtained at the 2-Hz rate. To qualitatively compare the data, color images of the sound-speed time series are used to show the behavior of the vertical sound-speed field. Figure 14 shows time series of the sound-speed field during the storm that peaks near year day 355, 1993 and whose mean depth profiles are shown in Fig. 11. The storm was chosen as representative of the typical response of the sound-speed field to increasingly rough conditions during a storm cycle. Each color image represents 40 min of data and panels are separated by approximately 4.5 h. At the onset of the storm, the reductions appear to be limited to the very near-surface region as shown in Fig. 14(a). As the seas develop, the scales of breaking increase, as displayed in the subsequent time series. The data shown in Fig. 14(b) demonstrate the effect of larger scale wave breaking injecting air to depth. Also apparent in the figure are the near-surface reductions that occur more frequently and persist for longer times. At the peak of the storm [Fig. 14(c)], the winds are in excess of 16 m/s and the significant wave heights approach 5 m. At this stage, the data reveal a persistent layer of reduced sound speed on which the deeper reductions are superimposed. Figure 14(b)–(e) show the response of the sound-speed field after the storm has peaked and begins to decay.

The large sound-speed reductions visible in Fig. 14(c), while variable in depth of penetration and residence time, appear to occur with some periodicity. To investigate the coupling of the wave field to the sound speed, the time series from the pitch–roll buoy are analyzed. The sea-surface elevation time series that corresponds with the sound-speed field shown in Fig. 14(c) is presented in Fig. 15(a). The low-frequency modulation of the wave field reflects the presence of wave groups or trains of high amplitude waves. A number of linear and nonlinear theories have been

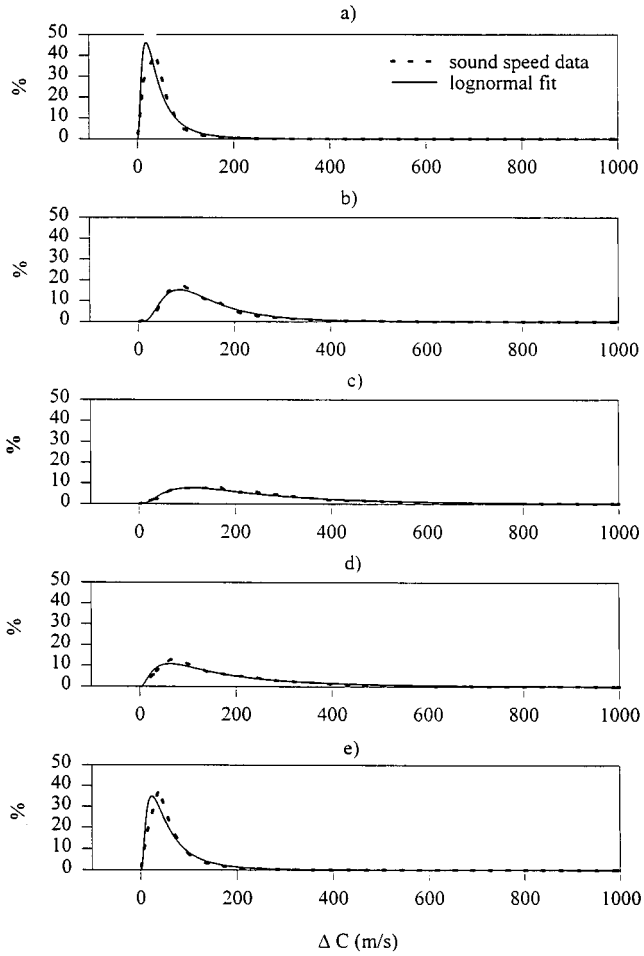


FIG. 17. Measured probability distributions (solid line) of the sound-speed reductions measured at 0.7 m during the storm shown in Fig. 11 and Fig. 14(a)–(e). A log-normal distribution (dotted line), based on the mean and variance of the signal is plotted with the data, providing good agreement over the course of the storm. Note the spreading of the distribution at the peak of the storm (c).

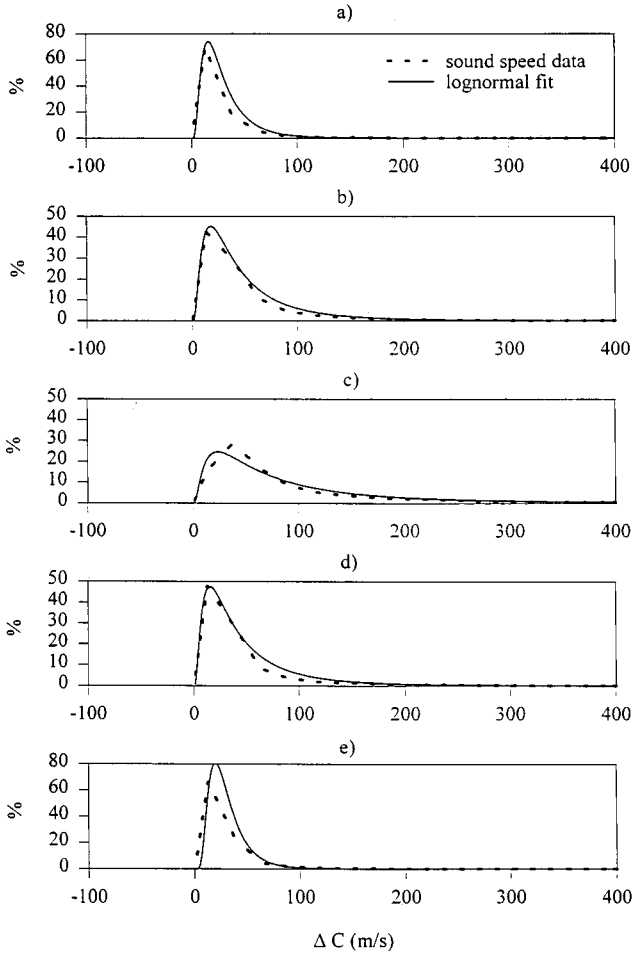


FIG. 18. Similar to Fig. 17(a)–(e) except the observed distributions are based on data obtained at the 1-m depth. Again, good agreement occurs between the predicted lognormal distribution (dotted line) and the data (solid line).

reviewed[44,45] that effectively describe the groupiness of the wave field. The envelope of the wave groups can be computed through application of the Hilbert transform.[46,47] If a complex time series $S(t)$ is formed

$$S(t)=\eta(t)+iH[\,\eta(t)\,], \tag{6}$$

where the surface displacement time series is denoted by $\eta(t)$ and $H[\,]$ is the Hilbert transform, the envelope of the displacement will be the magnitude of the complex signal, $|S(t)|$. This envelope is represented by the dashed line in Fig. 15(a). The corresponding power spectra of both the displacement and envelope signal are shown in Fig. 15(b), displaying distinct peaks for both the wind–wave spectrum $O(0.1)$ Hz and for the envelope spectrum at $O(0.01)$ Hz. A comparison between the envelope spectrum and the spectra of the near-surface sound-speed measurements obtained at the shallowest depths is shown in Fig. 16(a). The figure presents good agreement between the peak of the wave group spectrum and peaks in the sound-speed spectra at the lower group frequencies of $O(0.01)$ Hz. Other comparisons between the wave envelope spectrum and the sound-speed field produce similar results during the various storm events over the course of the experiment. Clearly the data suggests a correspondence between the groupiness of the wave field and the frequency of the large sound-speed reductions. This is consistent with aerial observations of the appearance of whitecaps by Donelan, Longuet-Higgins, and Turner.[48]

The depth of the injection events is evaluated through analysis of the response of the contour depth which represents the depth at which the sound speed is 90% of the bubble-free sound speed. This contour was selected (void fraction $\beta\cong 10^{-5}$) as its depth typically ranged from 1 to 3 m during these conditions ($U_{10}=16$ m/s, $H_{1/3}=5$ m). This depth range was chosen for the vertical resolution of the sound-speed module spacing. The spectrum of the depth of this layer is compared with the spectrum of the near-surface sound-speed measurement [Fig. 16(b)]. Agreement is evident between the two spectra at the lower wave group frequencies. The data indicates that the modulation of the penetration depth coincides with large sound-speed reduction events measured at the shallowest depth. Furthermore, the high levels of coherence between the near-surface sound speed and the contour depth [Fig. 16(c)], especially at the lower wave group frequencies, highlight the significance of the wave field in controlling not only the frequency of injection events, but the penetration of high bubble concentrations.

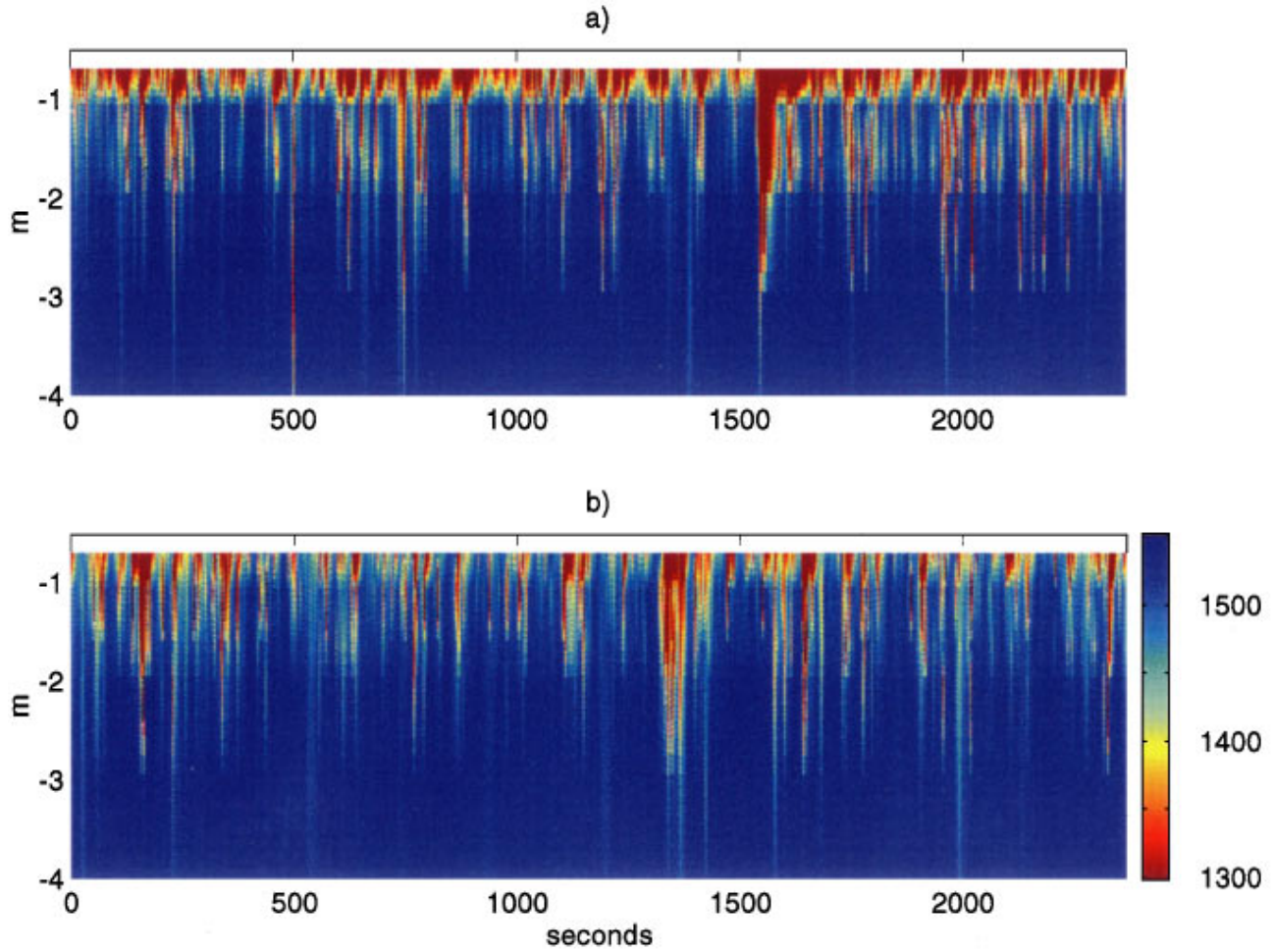


FIG. 19. Color contour plots of sound-speed time series during the peak of a storm occuring (a) 1993 year day 370 and (b) 1993 year day 380. While both time series exhibit the large, penetrating air injection events, the later storm (b) appears to lack a permanent layer of reduced sound speed just near the surface.

Figure 16(a) and (b) also shows evidence of a significant signal in the sound-speed and penetration depth at a frequency of 2×10^{-3} Hz. This does not appear to be associated with surface effects (although there is an unexplained peak in the envelope spectrum at 3×10^{-3} Hz) but may be associated with quasi-periodic Langmuir circulations. It is now well known that the lower-frequency modulation of acoustic backscatter measured by surface sonars at grazing incidence is dominated by scattering from bubbles entrained into Langmuir circulations.[49] The measurements reported here demonstrate that much of the variance associated with bubble-field parameters may be directly related to wave phenomena, rather than indirectly through Langmuir circulations, although both may play a role.

The probability distribution for the sound-speed reductions at 0.7- and 1-m measurement depths are shown in Figs. 17 and 18 for the previously discussed storm. The distributions are formed from the 40 min records using histograms with 25 m/s bins. Also shown are the corresponding log-normal distributions, based on the means and variances of the signals. During periods of intense wave breaking, the distribution spreads, indicating more occurrences of larger sound-speed reductions. At the beginning and the end of the storm, the distributions narrow, indicating fewer large reductions. Eventually, the distribution asymptotes to a delta function representing bubble-free water indicative of calm seas. Comparison between the two depths reveals that the deeper depth provides a narrower distribution, implying that the largest reductions occur closest to the surface. The trends in both the data and the calculated log-normal distributions appear to be in close agreement during periods of breaking throughout the 43-day record, and appear to extend to depths of approximately 3 m during high sea states. During calm periods, or at depths where there is little variation in the sound-speed signal, the agreement between the lognormal distribution and the data begin to deviate significantly.

A feature of the average data that deserves attention is the seemingly suppressed response of the sound-speed field to the storm that occurs near year day 380, 1993 (refer to Fig. 8). The suppression is evident when comparing the average sound speed during the storm to averages of earlier storms. A number of instrumentation issues were investigated in an attempt to determine if the differences were due to measurement errors. Effects of biofouling, low battery power, and buoy motion were considered, but provided no clear explanation for the apparent differences in the later storm cycle. In an effort to resolve the differences, color contours of the sound-speed field are created for the peaks of a number of storms. Figure 14(c) and Fig. 19(a) and (b) show the sound-speed field during storms under similar wind-

speeds. A common feature between the three storms is the presence of deep, intermittent sound-speed reductions that tend to persist for $O(30)$ s, similar to the reduction shown in Fig. 6. These are the reductions that appear to vary at frequencies similar to the group frequencies of the surface wave field. A visible difference between the storm late in the experiment, Fig. 19(b), and the earlier storms [Figs. 19(a) and 14(c)], is a diminished near-surface layer of reduced sound speed.

If the frequent, shallow injection of air is attributed to the breaking of smaller scale waves, the suppressed response of the sound-speed field to the storm of 1993 day 380 suggests a decline of these smaller breakers during this period. One explanation for the suppressed small scale breaking is the drop in the magnitude of the surface currents at this time [Fig. 7(c)]. The elevated surface currents earlier in the record are due to the passage of an eddy that spun off the nearby Gulf Stream. Since the magnitude of the current is the same order of magnitude as the phase speed of higher-frequency wind waves $[O(0.5–2)\text{Hz}]$, it is feasible that breaking of waves in these frequencies would be highly variable, dependent not only on the magnitude of the wind and current but the relative directions of the two. Observations have also shown that the penetration depths of bubble clouds are sensitive to stability of the surface waters.[1] In order to substantiate these effects, additional measurements would be required with an independent measurement of wave breaking. For example, the combination of video measurements of whitecap coverage and near-surface acoustic techniques, as well as measurements of water column stability, would provide the necessary data to further understand these processes.

VI. DISCUSSION AND CONCLUSION

This paper reports on the results from the ASREX field experiment during which direct sound-speed measurements were acquired over a wide range of oceanographic conditions. A specialized buoy was designed and built which provided an autonomous platform from which direct sound-speed measurements could be made for an extended period. The data provide insight into the highly variable sound-speed field.

The wind speed provides strong correlations with the mean and rms sound speeds for wind speeds greater than approximately 8 m/s at the measurement depth closest to the surface. At the measurement depth of 1.6 m, the wind speed correlation appears robust, extending to lower wind speeds. The differences between the two depths indicate the variability of the entrainment of air in the shallow surface layer. As expected, comparisons of the mean and rms sound speeds with the significant wave height produced significant scatter, supporting previous work that indicates that the significant wave height may not be the best indicator of wave breaking at the ocean surface.

A parameter measuring the steepness of the wave field is calculated from data obtained from a pitch–roll buoy and correlated with the mean and rms sound-speed values. The computed steepness parameter provides the basis for a considerably improved correlation compared with the correlations of wind speed and significant wave height. Further evidence for the dependence of air injection on the wave field is provided in the comparisons of the wave group frequencies and the frequency of sound-speed reductions. Analysis of the sound-speed time series reveals large injection events that last for O(30)s at periods similar to those which correspond to the wave groups. The depth of the contour which represents the depth at which the sound speed is 90% of the bubble-free sound speed is also found to modulate at similar frequencies. Lower-frequency modulations may be due to bubble transport by Langmuir circulations or low-frequency modulation of the wave envelope by distant swell.

It was found that Phillips[36] wave dissipation model, based on the wind and wave fields over the course of the experiment, gave significant correlations with both the mean and rms sound-speed measurements. It is anticipated that similar bulk dissipation estimates may be used for predicting other near-surface acoustic processes that are affected by wave breaking. Further refinements to the dissipation estimate, perhaps by inclusion of swell and current interactions, may better address the variability of the near-surface layer.

Despite contrasts in their measurement techniques and results, the previous field work by Lamarre and Melville[25] and Farmer and Vagle[22] found that an exponential function described the depth dependence of the sound-speed anomaly. The application of the exponential sound-speed depth profile is due in part to earlier work by Thorpe[1,2] and the fact that the low-frequency sound-speed reduction is proportional to void fraction for $\beta<10^{-5}$. Modeling the vertical transport of bubbles using a turbulent transport model, Thorpe obtains an exponential profile when the diffusion coefficient is assumed to be constant with depth. The suggestion of a power-law dependence, specifically an inverse square profile, was made by Buckingham[50] in his work on bubble-generated acoustic waveguides. While the inverse square profile was initially suggested as a means to provide agreement between mode-filtering theory and observations of ambient noise below bubble clouds, a more physically based explanation has recently been proposed by Buckingham.[51]

The ASREX data set suggests that power laws in the range -2 to -4 describe the depth profiles. Recent measurements in the near-surface layer[52,43] show that levels of turbulent dissipation are one to two orders of magnitude higher than levels predicted by the previously accepted ideas of a wall turbulent boundary layer. In light of the variability in the ASREX data set and recent dissipation measurements, it is clear that a better knowledge of the near-surface turbulence is needed to give better turbulent transport models of bubbles.

The bulk of the effort presented here has been in the analysis of the 3.33-kHz sound-speed data. This was due in part to the fact that the lowest frequency we measured would best approximate the low-frequency/nondispersive limit. However, a brief discussion of the sound-speed data from the other two frequencies is warranted. Preliminary analysis of the 5- and 10-kHz data over the storm cycle discussed earlier (year day 355, 1993), provides evidence that the dispersive effects of bubbles may extend to frequencies as low as 5 kHz. Figure 20 presents time series of the mean sound speeds measured at (b) 0.7 m and (c) 1.6 m for all three

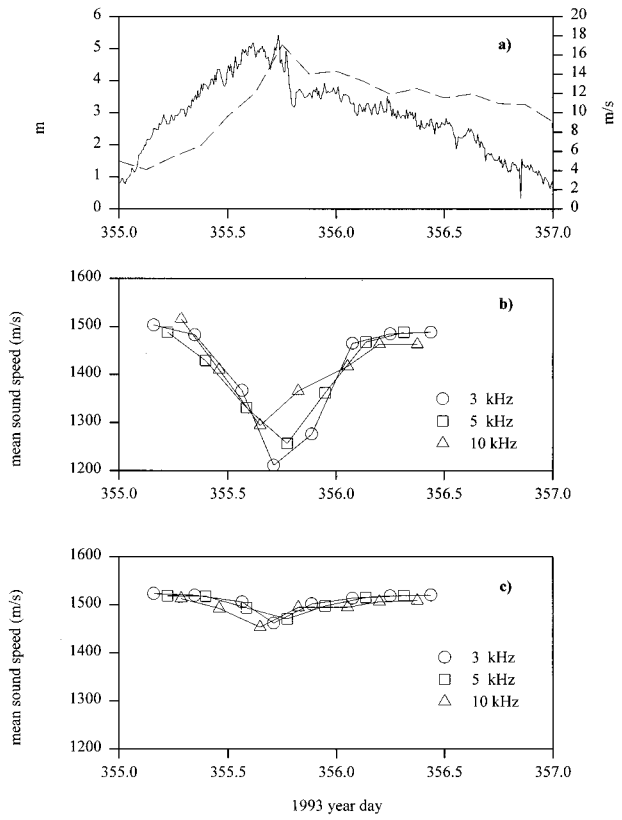


FIG. 20. Response of the sound-speed field over the course of a storm showing the effects of dispersion (year day 355, 1993). Times series show (a) significant wave height (dashed line) and wind speed (solid line) and the mean sound speed for 3.33, 5, and 10 kHz at the (b) 0.7-m and (c) 1.6-m measurement depths.

frequencies during the storm. While the measurements are not concurrent (a separation of 50 min between the end of one frequency and beginning of another), the means appear to track quite well together at the beginning of the storm. At the shallowest measurement, the means appear to deviate near the peak of the storm ($\sim$day 355.5). These deviations may be explained by a significant increase in bubbles of radii that resonate at or near 5 kHz ($a_r \approx 650\ \mu$m) during this time. Data from the deeper measurement do not appear to support this trend. However, it would be expected that fewer larger bubbles would exist at greater depths due to buoyancy effects and breakup mechanisms, resulting in the nondispersive limit shifting to higher frequencies.

ACKNOWLEDGMENTS

The authors are indebted to a number of people whose professionalism led to the success of the ASREX project. We thank the officers and crew of both the R/V KNORR and R/V EDWIN LINK for providing a safe working environment for the deployment and recovery of the various moorings. The success of the moorings is due to the design and deployment by John Bouthilette, John Kemp, George Tupper, and Bryan Way of WHOI. We thank Bob Weller and Anand Gnanadesikan from WHOI and GFDL for providing the wave and meteorological data and Russ Davis from SIO for loaning the buoy floatation and tower. This project was funded by ONR (ARSRP and Acoustics).

[1] S. A. Thorpe, ''On the clouds of bubbles formed by breaking waves in deep water and their role in air-sea gas transfer,'' Philos. Trans. R. Soc. London, Ser. A **304**, 155–210 (1982).
[2] S. A. Thorpe, ''The role of bubbles produced by breaking waves in supersaturating the near-surface ocean mixing layer with oxygen,'' Ann. Geophys. (France) **2**, 53–56 (1984).
[3] W. M. Carey and D. Browning, ''Low frequency ambient noise: measurements and theory,'' in *Sea Surface Sound*, edited by B. R. Kerman (Kluwer Academic, Dordrecht, 1988), pp. 361–376.
[4] H. Medwin and M. M. Beaky, ''Bubble sources in the Knudsen sea noise spectra,'' J. Acoust. Soc. Am. **86**, 1124–1130 (1989).
[5] H. Medwin and A. C. Daniel, ''Acoustical measurements of bubble production by spilling breakers,'' J. Acoust. Soc. Am. **88**, 812–40 (1990).
[6] A. Prosperetti, ''Bubble-related ambient noise in the ocean,'' J. Acoust. Soc. Am. **84**, 1042–1054 (1988).
[7] M. R. Loewen and W. K. Melville, ''An experimental investigation of the collective oscillations of bubble plumes entrained by breaking waves,'' J. Acoust. Soc. Am. **95**, 1329–1343 (1994).
[8] F. S. Henyey, ''Acoustic Scattering from ocean microbubble plumes in the 100 Hz to 2 kHz region,'' J. Acoust. Soc. Am. **90**, 399–405 (1991).
[9] P. M. Ogden and F. T. Erskine, ''An empirical prediction algorithm for low-frequency acoustic surface scattering strengths,'' Naval Research Laboratory, Washington DC, 1992.
[10] B. D. Johnson and R. C. Cooke, ''Bubble populations and spectra in coastal waters: a photographic approach,'' J. Geophys. Res. **84**, 3761–3766 (1979).
[11] A. L. Walsh and P. J. Mulhearn, ''Photographic measurements of bubble populations from breaking wind waves at sea,'' J. Geophys. Res. **92**, 14553–14565 (1987).
[12] M.-Y. Su, S. C. Ling, and J. Cartmill, ''Optical microbubble measurement in the North Sea,'' in *Sea Surface Sound*, edited by B. R. Kerman (Kluwer, Dordrecht, 1988).
[13] S. Baldy, ''Bubbles in the close vicinity of breaking waves: statistical characteristics of the generation and dispersion mechanism,'' J. Geophys. Res. **93**, 8239–8248 (1988).
[14] P. Geissler and B. Jahne, ''A 3D-sensor for the measurement of particle concentration from image sequences,'' in *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXI, Part B5, ''Invited and Presented Papers of the XVIIIth Congress,'' Vienna, Austria, 9–19 July 1996, edited by P. Waldh (published by the Committee of the XVIII International Congress for Photogrammetry and Remote Sensing).
[15] M. V. Hall, ''A comprehensive model of wind generated bubbles in the ocean and predictions of the effects on sound propagation at frequencies up to 40 kHz,'' J. Acoust. Soc. Am. **86**, 1103–1117 (1989).
[16] H. Medwin, ''In situ acoustic measurements of bubble populations in coastal waters,'' J. Geophys. Res. **75**, 599–611 (1970).
[17] H. Medwin, J. Fitzgerald, and G. Rautman, ''Acoustic miniprobing for ocean microstructure and bubbles,'' J. Geophys. Res. **80**, 405–413 (1975).
[18] H. Medwin, ''In situ acoustic measurements of microbubbles at sea,'' J. Geophys. Res. **82**, 971–976 (1971).
[19] H. Medwin and N. D. Breitz, ''Ambient and transient bubble spectral densities in quiescent seas and under spilling breakiers,'' J. Acoust. Soc. Am. **88**, 408–412 (1989).
[20] M.-Y. Su, D. Todoroff, and J. Cartmill, ''Laboratory comparisons of acoustic and optical sensors for microbubble measurement,'' J. Atmos. Oceanic Tech. **11**, 170–181 (1994).
[21] S. Vagle and D. Farmer, ''The measurement of bubble size distributions by acoustical backscatter,'' J. Atmos. Oceanic Tech. **9**, 630–644 (1991).
[22] D. Farmer and S. Vagle, ''Waveguide propagation of ambient sound in the ocean surface bubble layer,'' J. Acoust. Soc. Am. **86**, 1897–1908 (1989).
[23] A. B. Wood, *A Textbook of Sound* (Bell, London, 1941).
[24] C. S. Clay and H. Medwin, *Acoustical Oceanography* (Wiley, New York, 1977).
[25] E. Lamarre and W. K. Melville, ''Sound-speed measurements near the ocean surface,'' J. Acoust. Soc. Am. **96**, 3605–3616 (1994).
[26] E. Lamarre and W. K. Melville, ''Instrumentation for the measurement of sound speed near the ocean surface,'' J. Atmos. Oceanic Tech. **12**, 317–329 (1995).
[27] E. Silberman, ''Sound velocity and attenuation in bubbly mixtures measured in standing wave tubes,'' J. Acoust. Soc. Am. **29**, 925–933 (1957).
[28] E. Lamarre, ''An experimental study of air entrainment by breaking waves,'' Ph.D. thesis, MIT/WHOI Joint Program, 1993.
[29] S. F. Barstow, G. Ueland, H. E. Krogstad, and B. A. Fossum, ''The

wavescan 2nd generation directional wave buoy,'' IEEE J. Ocean Eng. **16**, 254–266 (1991).
[30] N. R. Galbraith, A. Gnanadesikan, W. M. Ostrum, E. A. Terray, B. S. Way, N. J. Williams, S. H. Hill, and E. Terrill, ''Meteorological and Oceanographic Data during the ASREX III Field Experiment: Cruise and Data Report,'' WHOI technical Report No. 96-10, Woods Hole Oceanographic Institution, Woods Hole, MA.
[31] H. Medwin, ''Speed of sound in water: a simple equation for realistic parameters,'' J. Acoust. Soc. Am. **58**, 1318–1319 (1975).
[32] F. Felizardo and W. K. Melville, ''Correlations between ambient noise and the ocean surface wave field,'' J. Phys. Oceanogr. **25**, 513–532 (1995).
[33] F. Felizardo, ''Ambient Noise and Surface Wave Dissipation in the Ocean,'' Ph.D. thesis, MIT/WHOI Joint Program, Cambridge/Woods Hole, MA, 1993, 219 pp.
[34] R. J. Rapp and W. K. Melville, ''Laboratory measurements of deep-water breaking waves,'' Philos. Trans. R. Soc. London, Ser. A **331**, 735–800 (1990).
[35] M. R. Loewen and W. K. Melville, ''Microwave backscatter and acoustic radiation from breaking waves,'' J. Fluid Mech. **224**, 601–623 (1991).
[36] O. M. Phillips, ''Spectral and statistical properties of the equilibrium range in wind-generated gravity waves,'' J. Fluid Mech. **156**, 505–531 (1985).
[37] K. Hasselmann, ''On the spectral dissipation of ocean waves due to white-capping,'' Boundary-Layer Met. **6**, 107–127 (1974).
[38] G. J. Komen, S. Hasselmann, and K. Hasselmann, ''On the existence of a fully developed wind-sea spectrum,'' J. Phys. Oceanogr. **13**, 816–827 (1984).
[39] W. K. Melville, M. R. Loewan, F. C. Felizardo, A. T. Jessup, and M. J. Buckingham, ''Acoustic and microwave signatures of breaking waves,'' Nature (London) **335**, 54–56 (1988).
[40] R. M. Kennedy, ''Acoustic radiation due to surface wave breaking,'' J. Acoust. Soc. Am. **94**, 2443–2445 (1993).
[41] G. J. Komen, L. Cavaleri, M. Donelan, K. Hasselmann, S. Hasselmann, and P. A. E. M. Janssen, *Dynamics and Modeling of Ocean Waves* (Cambridge U.P., Cambridge, England, 1994), 532 pp.
[42] W. K. Melville, ''Energy dissipation by breaking waves,'' J. Phys. Oceanogr. **24**, 2041–2049 (1994).
[43] W. K. Melville, ''The role of surface-wave breaking in air–sea interaction,'' Annu. Rev. Fluid Mech. **28**, 279–321 (1996).
[44] S. Elgar, R. T. Guza, and R. J. Seymour, ''Groups of waves in shallow water,'' J. Geophys. Res. **89**, 3623–3634 (1984).
[45] H. C. Yuen and B. M. Lake, ''Instabilities of waves on deepwater,'' Annu. Rev. Fluid Mech. **12**, 303–334 (1980).
[46] K. T. Shum and W. K. Melville, ''Estimates of the joint statistics of amplitudes and periods of ocean waves using an integral transform technique,'' J. Geophys. Res. **89**, 6467–6476 (1984).
[47] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures* (Wiley, New York, 1986).
[48] M. Donelan, M. S. Longuet-Higgins, and J. S. Turner, ''Periodicity in whitecaps,'' Nature (London) **239**, 449–450 (1972).
[49] A. J. Plueddemann, J. A. Smith, D. M. Farmer, R. A. Weller, W. R. Crawford, R. Pinkel, S. Vagle, and A. Gnanadesikan, ''Structure and variability of Langmuir circulation during the Surface Waves Processes Program,'' J. Geophys. Res. **101**, 3525–3543 (1996).
[50] M. J. Buckingham, ''On acoustic transmission in ocean-surface wave guides,'' Philos. Trans. R. Soc. London, Ser. A **335**, 513–555 (1991).
[51] M. J. Buckingham, ''Sound speed and void fraction profiles in the sea surface bubble layer,'' Appl. Acoust. **51**, 225–250 (1997).
[52] Y. C. Agrawal, E. A. Terray, M. A. Donelan, P. A. Hwang, A. J. Williams III, W. M. Drennan, K. K. Kahma, and S. A. Kitaigorodskii, ''Enhanced dissipation of kinetic energy beneath breaking waves,'' Nature (London) **359**, 219–220 (1992).

Benchmarking virtual backscatter in parabolic evolution

Eric Smith
Applied Research Laboratories, The University of Texas at Austin, P.O. Box 8029, Austin, Texas 78713-8029

(Received 2 October 1995; revised 16 October 1996; accepted 8 July 1997)

A perturbative method has recently been presented for improving parabolic equations to account for intermediate backscattering induced by slow range dependence of the acoustic index of refraction. In realistic cases, it yields solutions to the full wave equation from a truly one-way algorithm, and generalizes a number of special results of this kind which have been known for many years. Numerically validating the method, though, involves certain unique difficulties, because the physical corrections it introduces to local evolution are accompanied by medium dependence of the projection operator that defines the input and output wave fields. A method for benchmarking is presented here, which permits numerical isolation of range-local effects from the endpoint projections, and also from particular details of the choice of one-way model used. To illustrate that the physical content of the new terms can be sensibly validated in this manner, an exact solution is considered, where they are shown to correct the ''idealized'' one way equation on which current parabolic algorithms are based. This solution is used to illustrate the relation of energy conservation to one-way evolution in the presence of slow range dependence. In particular, it demonstrates that energy conservation *alone* is not a sufficient requirement to produce correct parabolic evolution, even when it pertains. © *1997 Acoustical Society of America.* [S0001-4966(97)00711-X]

PACS numbers: 43.30.Ft, 43.30.Gv [JHM]

INTRODUCTION

Parabolic equations[1] represent an attempt to factor the quadratic Helmholtz equation into a pair of linear evolution equations. Because evolution involves only a first derivative along some ''range'' coordinate, it can be implemented as a forward marching algorithm, which is noniterative and therefore highly efficient.[2] The term parabolic equation (PE) originally denoted a lowest-order factorization in powers of the transverse Laplacian, suitable for describing narrow-angle propagation. Nowadays, it is used to refer to any of a number of wide- as well as narrow-angle algorithms, which seek to approximate an ''idealized one-way wave equation,'' in which the evolution operator is defined formally as the square root of the nonrange-derivative part of the Helmholtz equation.[3]

Generating useful wave field models with PEs involves two difficulties. The first is accurately and efficiently implementing a numerical approximation to the square root operator, an area in which much progress has been made.[4,5] For range-independent systems, implementation is the only issue, because the idealized one-way equation and its Hermitian conjugate give an exact factorization of the Helmholtz equation. In range-dependent systems that is no longer the case, and any implementation based on the same evolution operator can at best converge to an approximation of the full-wave solution. Much less of a general nature has been done prior to now to overcome this problem, so the square root operator of Ref. 3 is usually taken as the best generally applicable starting point to define PEs, even for range-dependent systems.

The essential incompatibility between range dependence and marching comes from multiple backscatter. If only the forward-going solution is of interest on the final range slice, one-time backscatter can be incorporated as a local correction in a marching algorithm, and at the same level of approximation, the corresponding reflected wave may be generated if desired.[6] Multiple backscatter, on the other hand, ultimately transports some energy forward in range, in a way that cannot be predicted from any range-local criterion. Even in systems with simple boundary conditions, changes in sound speed along the propagation path can create coupling effects that are not simple.

A class of special cases is known in which full-wave solutions in range-dependent environments can be obtained with one-way methods. These are the cases where a range-dependent problem can be related by conformal transformation to one that is range independent.[7,8] Full range independence of the mapped system is not even strictly required, because conformal transformations can be used to replace changes in sound speed with changes in geometry (hence boundary position), many forms of which can be handled effectively using current PE methods.[9]

Unfortunately, in three dimensions, even slowly varying sound speed profiles are not generally conformal to isospeed systems. Further, while by theorem they always are in two dimensions, because of the much larger symmetry of the two-dimensional conformal group, finding the required transformation may be as difficult as solving the wave equation directly. This method is thus not practical for modeling, although it can be used to efficiently generate benchmarks,[8,2] suitable even for testing coupled-mode codes.[10] No adequate answer has therefore been extracted, from these methods alone, to the question ''What is the most general case for which a solution generated by marching can converge to a solution of the full wave equation, and how can such an algorithm be implemented?''

Recently, a perturbative factorization of the Helmholtz equation was presented,[11] based on the Foldy–Wouthuysen (FW) transformation[12] from the early days of relativistic classical field theory.[13] It is related to both the usual scalar factorization and the concept of the splitting matrix,[14,15] which reduces the Helmholtz equation exactly to a pair of linear, but coupled evolution equations. In particular, in range-independent cases it reduces to a power-series representation of the optimal splitting matrix of Ref. 15, with evolution determined by the standard idealized one-way wave equation. The value of FW is that it preserves the formal exactness of the splitting-matrix method in a useful class of range-dependent cases, while generalizing the usual one-way equation in a manner consistent with PE implementation. As a result, it provides an answer to the question previously posed, and generalizes the known special cases obtained by conformal transformation to eliminate the requirement of unphysically large symmetry.

The FW procedure applies whenever energy is conserved *asymptotically* in the linearly independent forward- and backward-going solutions to the full-wave equation. In this context, it gives (by explicit construction) the relation of energy conservation to one-way evolution. In particular, it can be used to show that energy conservation can be consistent with one-way evolution, but still not constitute an unambigious constraint to define that evolution, although it has been used as one.[16]

To put this conclusion in perspective, one may consider the problem of large-scale acoustic ocean tomography.[17] Model solutions must be able to predict the evolution of pulses that propagate over megameter scales, through very complex environments. Keeping computational cost to a minimum in such cases is essential. One of the most difficult aspects of pulse evolution to correctly predict in varying sound speed environments, though, is that of the *phase correlations* among the different frequencies that make up the pulse. As is well-known, current energy conserving but range-dependent one-way algorithms produce their most significant errors in phase,[18] because of the omission of the so-called ''commutator term'' of the range derivative with the range-varying evolution operator. The omitted terms are (on dimensional grounds) scaled by inverse powers of frequency, so the phase errors are necessarily frequency-dependent, corrupting phase correlations as well as introducing overall offsets.

It is important to recognize, though, that such propagation is *itself* only possible because the physical systems of interest are special in a different way. After large short-range losses, acoustic energy is efficiently preserved within water column ducts or channels. For the most part, changes in sound speed occur gradually enough along the propagation path that they do not even direct energy out of the water column. The gradual nature of variations in these systems makes them ''simple'' in precisely the correct way for FW to apply.

Benchmarking an implementation of the FW procedure introduces some unique problems, though, associated with the form it gives to the splitting matrix. This matrix, which relates the linearly independent components that evolve under a one-way equation to full-wave solutions, must be controlled as an operator expansion together with the resulting PE, introducing a potential source of propagating errors not related to the range-local, physical FW corrections. Unlike the splitting matrices considered for practical implementation in Ref. 15, it involves medium-dependent normalization terms, so it is not only needed to relate the forward- to the backward-going component; the explicit form must be taken into account to obtain the shape of the full-wave solution.

In view of these facts, the purpose of this paper is threefold:

(i) To show that the new effects introduced by FW, even at perturbative order, are corrections to the usual *idealized* one-way wave equation, independent of other approximations that may be introduced in implementing it.

(ii) To provide a method for benchmarking these corrections in the context of a single solution algorithm, such as a conventional PE. This bootstrap method makes it possible to separate analysis of the physically relevant changes to the one-way evolution operator from other issues of implementation, such as type of PE algorithm used or order of numerical approximation.

(iii) To illustrate the physical significance of the perturbative FW corrections, their relation to energy conservation, and the kind of error that arises if the latter is applied incorrectly as a constraint.

The organization is as follows: To establish notation, the scalar factorization of the Helmholtz equation is reviewed, and the effects that can be created by a range-dependent sound speed profile are categorized in Sec. I. The proposed benchmarking method will be based on conformal transformation, and this is briefly reviewed in Sec. II. To demonstrate the meaning and correctness of the FW perturbations in a context independent of numerics, an exact solution will be considered. This is introduced, and its relevant properties are derived in Sec. III. In Sec. IV, the FW diagonalization procedure is defined and applied, both perturbatively and exactly, to the analytic solution of the previous section. Criteria associated with the choice of small-parameter expansion used are also mentioned. The relation of these solutions to energy conservation is examined in Sec. V, Sec. VI gives examples, and Sec. VII presents a discussion of the results in the terms of an effective acoustic index of refraction.

I. ONE-WAY EVOLUTION

It will be assumed that the desired field is the pressure of a sound wave in some fluid, in a time-independent environment. Isolated frequency components may therefore be considered, of the form

$$P(\mathbf{x},t)=\mathrm{Re}[e^{-i\omega t}\phi(\mathbf{x})]. \tag{1}$$

The position coordinate is divided into range and transverse coordinates as $\mathbf{x}=(x,\mathbf{d})$, with the transverse part of the gradient denoted ∇. If the density may be regarded as constant, the field ϕ satisfies the Helmholtz equation,

$$\left(\frac{\partial^2}{\partial x^2}+\nabla^2+k_0^2n^2\right)\phi=0, \tag{2}$$

where $k_0 \equiv \omega/c_0$, and c_0 is some reference sound speed. The acoustic index of refraction, $n(\mathbf{x})$, relates the reference speed to the local value of the sound speed $c(\mathbf{x}) = c_0/n(\mathbf{x})$.

If the index $n(\mathbf{x})$ had a constant value of 1, Eq. (2) could be factored to yield

$$\left(-i\,\frac{\partial}{\partial x} + \sqrt{k_0^2 + \nabla^2}\right)\left(i\,\frac{\partial}{\partial x} + \sqrt{k_0^2 + \nabla^2}\right)\phi = 0. \tag{3}$$

These factors define the idealized one-way equations in the range- and depth-independent cases.[19] One of them must project out any term in the full solution, which therefore decomposes into forward- and backward-moving components that evolve separately. When $n(\mathbf{x})$ is not constant, a simple substitution of $k_0 n(\mathbf{x})$ for k_0 in Eq. (3), as it appears in Eq. (2), no longer factors the full-wave equation. The three qualitatively different types of coupling between forward- and backward-going modes that can be created by range varying $n(\mathbf{x})$ are:

(1) Sharp gradients in a sound speed profile, in which the fractional changes are large on the scale of a wavelength, can create real reflections of sound that transport energy back, asymptotically, in the direction from which it came. The process is essentially like reflection from a boundary, and is the case considered in Ref. 6. As noted, they are not usually a major influence on long-range propagation in ocean acoustics, so there are broad classes of systems in which it is legitimate not to consider them.

(2) Even very gradual changes in sound speed can couple forward- and backward-propagating waves. Sometimes this coupling results in no net backscattering of energy to asymptotic regions, but affects the forward propagation nonetheless. These intermediate backscatterings from gradual variations are referred to as *virtual*, to distinguish them from secondary scatterings (therefore also intermediate) that can occur in the course of real (asymptotic) backscatter as well. Only virtual backscatter occurs, whenever the scale of variations is much longer than the wavelength.

(3) By means of the FW method introduced in Ref. 11, virtual backscatter can be calculated perturbatively in small sound speed gradients, and is an energy-conserving process. However, the perturbative calculation may fail to converge, or it may converge only in the sense of an asymptotic expansion, missing effects that cannot be expanded in a power series about zero coupling. An example of the former case is refraction of a mostly sideways moving wave backward in range. This occurs even in the high-frequency geometric acoustics limit, and induces a breakdown of parabolic evolution if one tries to propagate such a wave past the turning point. An example of the latter is ''tunneling'' through stratified sound speed barriers, as has been calculated in WKB approximation.[20] Usually an outright failure of perturbation theory can be averted by choosing to evolve in an appropriate direction. Nonperturbative methods to compute energy-conserving backscatter fall outside the domain of this paper.

The distinction of interest is not whether mode coupling by sound speed gradients occurs; in realistic systems, that is generic. Rather, it is whether the absence of *asymptotic* backscatter makes it possible to divide full-wave solutions into separate forward- and backward-propagating parts. In the latter case, it is easy to see what is wrong with the standard one-way equation, and also why it *should* be possible to fix it.

In some small neighborhood of any point, a coordinate choice could always be made so that evolution was perpendicular to the gradient of $n(\mathbf{x})$ at that point. In that case, the standard one-way equation would be locally an exact factorization of Eq. (2), and refractive effects would be governed by transverse derivatives of $n(\mathbf{x})$ in the expansion of Eq. (3). In more general coordinates, the same derivative terms would not be represented by transverse gradients alone, but but would require range derivatives of $n(\mathbf{x})$ to appear as well. The physical refractive effects are the same, though, so such an inclusion in evolving the same scalar component should be possible.

The way such terms must be incorporated is subtle, though. This can be seen because, alternatively, local coordinates could be chosen so that evolution was *along* the gradient of $n(\mathbf{x})$. In Eq. (2), $n(\mathbf{x})$ acts like a *dilation* factor relating the frequency to the local scale for wavelengths. As such, it defines a *physical* geometry, seen by the propagating sound wave, relative to the coordinates $(x, \mathbf{d})$. This geometry is generally curved. A real solution, propagating in such a curved geometry, is diluted or focused and changes in magnitude. For evolution along a gradient, as will be seen, the physical coordinates look like angular coordinates in a wedge, and the correct solution should look locally like some wedge mode.[21,22] The same interference effects that result in geometric dilution of wedge modes necessarily also affect the phase advance. It will be seen that the latter effect can be incorporated by effectively shifting $n(\mathbf{x})$ by terms proportional to its range derivatives. The former, however, involves changes in amplitude, which are not produced by a Hermitian evolution operator alone. This is why one-way evolution that correctly captures all refractive effects requires a splitting matrix that involves normalization terms, to relate the one-way component to the full solution. These points will be seen explicitly in Sec. IV.

Thus Eq. (3) does not generalize to nonconstant $n(\mathbf{x})$ because, if the form of the operator is kept the same, it does not represent the same evolution in two different choices of coordinates. If an individual back- or forward-moving component of the full-wave solution does conserve energy asymptotically, then *it* makes no reference to such a distinction. The requirement of coordinate independence could be seen as the fundamental principle that defines the effective acoustic index. Therefore it is natural that many of the known special cases are derived with this tool.

II. CONFORMALLY EQUIVALENT SYSTEMS

Conformal coordinate transformations are most easily handled as analytic mappings of the complex plane. We define range and depth coordinates (x,y), respectively, and consider systems symmetric in the other transverse direction. Then, in terms of complex coordinates $z = (x+iy)/2$, $\overline{z} = (x-iy)/2$, the wave equation (2), for an isospeed system, takes the form

$$\partial_z \partial_{\overline{z}} \phi + k_0^2 \phi = 0, \tag{4}$$

where $\partial_z \equiv \partial/\partial z$ and $\partial_{\overline{z}} \equiv \partial/\partial \overline{z}$. It is convenient, for changing coordinates, to obtain Eq. (4) as the equation of motion from an action

$$S = \int dz\ d\overline{z}\ \frac{1}{2}\left(\partial_z \phi\ \partial_{\overline{z}} \phi - k_0^2 \phi^2\right). \tag{5}$$

New coordinates related to (x,y) by a conformal transformation are any functions (x',y') for which the complex variables $w=(x'+iy')/2$, $\overline{w}=(x'-iy')/2$ are, respectively, analytic and anti-analytic functions $w(z)$, $\overline{w(z)}$. By a change of variables, the same action in the new coordinates becomes

$$S = \int dw\ d\overline{w}\ \frac{1}{2}\left(\partial_w \phi\ \partial_{\overline{w}} \phi - k_0^2 \left|\frac{dz}{dw}\right|^2 \phi^2\right), \tag{6}$$

from which the corresponding equation of motion is now

$$\partial_w \partial_{\overline{w}} \phi + k_0^2 n^2 \phi = 0, \tag{7}$$

where $n^2 = |dz/dw|^2$. The constant index of refraction $n=1$ in Eq. (4) has been replaced with one depending on the coordinates w, $\overline{w}$, but the form of the Helmholtz equation has remained unchanged. By running the same process in reverse, a conformal transformation can be used to replace a varying sound speed with one which is constant.

If the system of interest as a benchmark is a waveguide with simple boundaries but varying sound speed, a conformal transformation can be used to remove the variation in sound speed, and the resulting change of coordinates tells where to place the new boundaries. Because the conformal transformations preserve angles, the derivative normal to a boundary in the old coordinates maps to a derivative which is normal to the boundary in the new ones as well. Thus any combinations of pressure release or rigid boundaries can be readily imposed on solutions.

The Helmholtz equation may be solved by any method which properly incorporates the boundary conditions, in the isospeed case, to produce a solution $\phi(z,\overline{z})$. This then provides field values in terms of the inverse transformation, $\phi(w,\overline{w}) = \phi(z,\overline{z})|_{z(w),\overline{z(w)}}$, which are also solutions of the wave equation in the varying index $n(w,\overline{w})$. These constitute the benchmark solution for a PE applied in those coordinates directly.

An exact solution generated with these methods, the case of the wedge itself, will demonstrate how the physical arguments of the last section are borne out in the FW solution. It will also show that the phase-correlation effects, correctly handled by FW but not by the standard one-way equation, involve local modifications to the evolution operator, in the sense of Bellman and Kalaba.[23] The amplitude effects depend on where the beginning and end range slices are chosen, and thus have no immediate local relation to properies of the sound speed at interior points. The key to defining a benchmark process that separates the two will be choosing conformal transformations where the endoint projections reduce to the constant forms recommended for practical application in Refs. 14 and 15.

III. EXACT SOLUTIONS

The conformal mapping of interest is given by $w(z) = (1/2\alpha)\log(2\alpha z)$. Here z is taken to have dimensions of a physical length, so α has the dimensions of a wave number. If z is written $z = \frac{1}{2} r\ e^{i\theta}$, this maps any wedge $\theta \in (\theta_-, \theta_+)$ to the strip $w = (x+iy)/2$, with $y \in ((1/\alpha)\theta_-, (1/\alpha)\theta_+)$. The origin of r is formally mapped to range $-\infty$ in the (x,y) strip. The corresponding index of refraction is given by $n(w,\overline{w}) = e^{\alpha(w+\overline{w})} = e^{\alpha x}$. The scale factor α is thus the log derivative $\alpha = \partial \log n/\partial x$. Exact solutions of Eq. (2) in both systems will be possible for any value of k_0, but the appropriate limit of slowly varying sound speed, for comparison to perturbation theory, will correspond to a choice with α/k_0 small.

Radial solutions of the Helmholtz equation in the (r,θ) coordinates are simply Bessel functions. Because the origin is mapped to a boundary ''point at infinity'' in (x,y), one may consider purely incoming or outgoing Green's functions centered at the origin $r=0$ as well. These Hankel function solutions correspond to purely forward- or backward-going solutions in the strip. In the wedge, the exact solutions are given by[24]

$$\phi_\nu^{\{\pm\pm\}}(z,\overline{z}) = H_\nu^{\pm}(k_0 r) e^{\pm i\nu\theta}. \tag{8}$$

Changing coordinates to the strip, these become the functions

$$\phi_\nu^{\{\pm\pm\}}(x,y) = H_\nu^{\pm}\left(\frac{k_0}{\alpha}\, e^{\alpha x}\right) e^{\pm i\nu\alpha y}, \tag{9}$$

which may be checked explicitly to solve the wave equation (2).

Solutions of this form have already been considered in Ref. 8. The concern in that work was with providing a modal expansion for the Green's function of a source in the strip. Here, the concern will be the phase and amplitude evolution of the individual Hankel functions themselves, so solutions of the wave equation in the strip will suffice. Of particular interest will be solutions involving Hankel functions of order 1/2, which have the closed form expression

$$H_{1/2}^{\pm}\left(\frac{k_0}{\alpha}\, e^{\alpha x}\right) = \sqrt{\frac{2\alpha}{\pi k_0}}\, e^{-(\alpha x/2) \pm i(k_0/\alpha)e^{\alpha x}}. \tag{10}$$

One may consider the wedge, and hence the strip, as having boundaries which specify particular allowed values of the angular frequency ν, or continue the (r,θ) plane indefinitely in θ and admit arbitrary ν.

By construction, as Green's functions in the wedge, energy flux carried out through circles of constant r is independent of r. The energy out per unit surface is defined in terms of the radial translation operator $-i\partial/\partial r$, and the total energy flux is the integral over that surface given by[25]

$$E \propto \int r\ d\theta\ \frac{1}{2}\left\{(i\partial_r \phi^\dagger)\phi + \phi^\dagger(-i\partial_r \phi)\right\}. \tag{11}$$

Because $r\ \partial/\partial r = \partial/\partial \log r = \partial/\partial(\alpha x)$, and $d\theta = \alpha\ dy$, this is the same as the expression for energy through a given range slice in the (x,y) system,

$$E \propto \int dy\, \frac{1}{2}\{(i\partial_x \phi^\dagger)\phi + \phi^\dagger(-i\partial_x \phi)\}. \tag{12}$$

Therefore, in both physical systems, an isospeed wedge and a strip with exponential sound speed profile, the forward- and backward-going Hankel function solutions conserve energy independently.

For the opposite-going basis functions $\phi^\pm$, one may group the terms that appear in the range derivative to write $-i\partial_x \phi^\pm = (i\mathcal{N} \pm \mathcal{H})\phi^\pm$, where $\mathcal{N}$ and $\mathcal{H}$ are Hermitian differential operators, and the $\pm$ reference to the angular dependence (θ or y) has been suppressed.[26] In this case, the fact that $\partial_y(n^2) \equiv 0$ implies that all of these operators and any of their derivatives commute. Substitution of these forms into Eq. (12) and use of Hermiticity gives the energy flux for the functions $\phi^\pm$ as

$$E^\pm \propto \pm \int dy\{\phi^{\pm\dagger}\mathcal{H}\phi^\pm\}. \tag{13}$$

Subsequent evaluation of the range derivative of $E^\pm$, again using Hermiticity, gives

$$\frac{\partial E^\pm}{\partial x} \propto \pm \int dy \left\{ \phi^{\pm\dagger}\mathcal{H}\left(\frac{\partial \log \mathcal{H}}{\partial x} - 2\mathcal{N} \right)\phi^\pm \right\}. \tag{14}$$

Thus conservation of energy flux for the two independent solutions determines that $\mathcal{N} = \frac{1}{2}(\partial/\partial x)\log \mathcal{H}$. Therefore, for these functions,

$$-i\,\frac{\partial}{\partial x}\,\phi^\pm = \left(\frac{i}{2}\,\frac{\partial}{\partial x}\,(\log \mathcal{H}) \pm \mathcal{H} \right)\phi^\pm. \tag{15}$$

For general ν, an asymptotic expansion for the Hankel functions is[27]

$$H_\nu^\pm(k_0 r) = \sqrt{\frac{2}{\pi k_0 r}}\, e^{\pm i(k_0 r - \nu(\pi/2) - \pi/4)} \times \sum_{n=0}^{N-1} \frac{(\mp)^n}{(2ik_0 r)^n n!}\,\frac{\Gamma(\nu + n + 1/2)}{\Gamma(\nu - n + 1/2)} + \mathcal{R}^N\left(\frac{1}{k_0 r}\right), \tag{16}$$

in which, in the (x,y) coordinates, $\mathcal{R}^N$ will be a remainder of order N in the small parameter α/k_0. From this expression, one can solve for $\mathcal{H}$ to leading nontrivial order in α/k_0 (which will be sufficient here) to obtain

$$\mathcal{H} = \sqrt{k_0^2 n^2 + \alpha^2(\tfrac{1}{4} - \nu^2)} + \cdots = \sqrt{k_0^2 n^2 + \alpha^2/4 + \partial_y^2} + \cdots\ . \tag{17}$$

The terms listed in Eq. (17) represent exact forms when $\nu = 1/2$. These expressions for the Hermitian part of the evolution operator will be useful for identifying various physical effects in what follows. Through Eq. (15) they also control the geometric, or alternatively refractive, dilution that conserves energy as the Hankel functions propagate. Because transverse derivatives of the sound speed profile vanish, that relation is particularly simple for these special cases. Establishing the precise relation in the general case, between the object which evolves according to an Hermitian one-way operator and the full solution that conserves energy and obeys the wave equation, is the the domain of the Foldy–Wouthuysen procedure, which was presented in Ref. 11.

IV. THE FOLDY–WOUTHUYSEN DIAGONALIZATION

A quadratic equation like (2) creates a different boundary-value problem than a linear evolution equation, because to evolve the full solution forward from a slice, the wave values and first range derivatives must be used. Linear evolution is defined entirely by the *values* of a wave field at a particular range slice, and thus must be governed by an operator expansion in transverse derivatives only. Replacing the initial-value derivative condition with a projection that selects the forward- and backward-moving components is the function of the splitting matrix. (The grouping of terms used here differs from Refs. 14 and 15, in expressing all such splitting explicitly as the effect of projection operators that act on scalar solutions of the Helmholtz equation.)

The first step in the FW procedure is to introduce a projection, **P**, involving only the zeroth and first range derivatives, such as

$$\mathbf{P}\phi \equiv \frac{1}{2}\begin{bmatrix} 1 - \dfrac{i}{k_0}\dfrac{\partial}{\partial x} \\ 1 + \dfrac{i}{k_0}\dfrac{\partial}{\partial x} \end{bmatrix}\phi \equiv \Phi. \tag{18}$$

P (equivalent to the splitting matrix T_0 of Ref. 15) extracts from any full solution ϕ of the wave equation two components, constructed to approximate the forward- and backward-moving parts of the full solution separately. The full solution can be recovered because the dot product $[1\ 1]\cdot\mathbf{P}$ is required to be 1. Whenever ϕ satisfies the full Helmholtz equation Φ satisfies a matrix equation of the form[28]

$$-i\,\frac{\partial}{\partial x}\,\Phi = \hat{H}\Phi. \tag{19}$$

For this projection, elementary algebra shows that the matrix $\hat{H}$ takes the specific form

$$\frac{\hat{H}}{k_0} = \left(\frac{\nabla^2}{k_0^2} + n^2 - 1 \right)\left(\frac{\eta + \rho}{2} \right) + \eta, \tag{20}$$

where the matrices η, ρ are defined to be

$$\eta = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \rho = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{21}$$

For a medium-independent projection **P**, such as the one chosen, the resulting $\hat{H}$ is not diagonal, which says that the components of Φ are not really purely forward- or backward-moving, but are mixed by the evolution operator $-i\partial/\partial x$. However, if the approximation is sufficiently close that the terms multiplying $(\eta+\rho)/2$ are all small, the dominant term is the remaining η, and at lowest order, the upper and lower components evolve with the correct phases.

As explained in Ref. 11, however, one can systematically obtain an Hermitian, diagonal evolution operator asymptotically in perturbation theory. The idea is to effectively

alter the projection operator **P** by considering, not Φ produced by the medium-independent splitting, but rather a matrix rotation $U\Phi$, which then evolves as

$$-i\,\frac{\partial}{\partial x}\,(U\Phi)=\widetilde{H}(U\Phi), \tag{22}$$

in which case the matrix $\widetilde{H}$ is determined by $\hat{H}$ and the choice of the rotation matrix U.

It is desirable to restrict U to be an element of the Lie group[29] generated by the matrices $\eta/2$, $\rho/2$, and $(i\eta\rho)/2$. It follows from the definition of this group (as shown in Ref. 11, where this is referred to as the ''pseudo-Unitary group'') that the components of $U\Phi$ describe the same total down range propagation of energy as the components of Φ. Because the pseudo-Unitary transformations are a group, the rotation U can be inverted, and Φ, and hence the full solution ϕ, recovered from $U\Phi$ after it has been evolved down range with $\widetilde{H}$. If $\widetilde{H}$ is proportional to the matrix η by some Hermitian differential operator, then by inspection the upper and lower components of $U\Phi$ evolve down range according to the two factors generalizing the idealized scalar one-way equation. Performing a rotation that accomplishes this is equivalent to obtaining a ''factorization'' of the Helmholtz equation that includes the commutator term. The construction of the matrix U gives the precise relation between the conjugate wave fields (the components of $U\Phi$) that evolve under a general factorization of Eq. (2), and the scalar full-wave solutions to the same equation.

To see that these effects, even though derived perturbatively, are corrections to the usual idealized one-way equation, and also that they cannot be determined from energy conservation alone, it is sufficient to apply this analysis to the exact solutions derived in the last section, and work up to leading order nontrivial corrections in α/k_0.

The first small operator expansion that will be considered is the one assumed in Ref. 11. It is a convenient expansion to use in introducing the method, because the assumed perturbative small parameter has a simple expression in the operator $\hat{H}$ of Eq. (20). The definitions presented here apply to the general case of transverse inhomogeneity, but the algebra will be greatly simplified because a number of commutators, involving transverse as well as range derivatives of the sound speed profile, vanish by symmetry when applied to the exact solution. They will therefore also be dropped in the notation.

The perturbative method is to construct the matrix U in stages. The original matrix $\hat{H}$ of Eq. (20) can be written $\hat{H}/k_0\equiv A^0\eta+B^0\rho+C^0(i\eta\rho)$. Defining an operator $X\equiv\frac{1}{2}(\nabla^2/k_0^2+n^2-1)$, it follows that $A^0=(1+X)$, $B^0=X$, $C^0=0$. If the expansion is performed about the position for which $n=1$ (which amounts to choosing the reference speed c_0), X is of order α/k_0.

One then approximates U by a sequence of products $U\approx U^N\cdots U^{\mathrm{III}}U^{\mathrm{II}}U^{\mathrm{I}}$, for which

$$-i\,\frac{\partial}{\partial x}\,(U^N\cdots U^{\mathrm{III}}U^{\mathrm{II}}U^{\mathrm{I}}\Phi)=\hat{H}^N(U^N\cdots U^{\mathrm{III}}U^{\mathrm{II}}U^{\mathrm{I}}\Phi). \tag{23}$$

The resulting sequence of $\hat{H}^N$ then converges to $\widetilde{H}$. The fact that, at each perturbative order, the next rotation performs a mixing of the previous order's approximate forward- and backward-moving components, says that the process is equivalent to including intermediate backscattering effects in new terms in the evolution operator. The matrix representation above for $\hat{H}$ was chosen because, with U^N elements of the pseudo-Unitary group, the successive approximations to $\widetilde{H}$ may be written in an invariant form, $\hat{H}^N/k_0\equiv A^N\eta+B^N\rho+C^N(i\eta\rho)$.

The first rotation is then taken to be $U^{\mathrm{I}}=e^{(X\eta\rho/2)}$ (to remove B^0), which produces

$$A^{\mathrm{I}}=\left\{1+X-\frac{X^2}{2}+\frac{X^3}{2}-\frac{X^4}{8}+\cdots\right\},$$
$$B^{\mathrm{I}}=\left\{-X^2+\frac{X^3}{3}+\cdots\right\},\quad C^{\mathrm{I}}=\left\{-\frac{\dot{X}}{2}\right\}, \tag{24}$$

where the overdot denotes the dimensionless range derivative $(1/k_0)(\partial/\partial x)$. Then, the choice $U^{\mathrm{II}}=e^{(-X^2\eta\rho/2-i\dot{X}\rho/4)}$ (obtained from B^1 and C^1) gives

$$A^{\mathrm{II}}=\left\{1+X-\frac{X^2}{2}+\frac{X^3}{2}-\frac{5X^4}{8}-\frac{\dot{X}^2}{8}+\cdots\right\},$$
$$B^{\mathrm{II}}=\left\{\frac{4X^3}{3}-\frac{\ddot{X}}{4}+\cdots\right\},\quad C^{\mathrm{II}}=\left\{\frac{3X\dot{X}}{2}+\cdots\right\}. \tag{25}$$

As the off-diagonal terms become higher order in X and range derivatives, the first deviation of the diagonal term from the expression for the square root operator $\sqrt{n^2+\nabla^2/k_0^2}$ appears as $-\dot{X}^2/8$, which for this sound speed profile looks like $-\alpha^2/8k_0^2$.

From the form (15) for the evolution of the Hankel function, the same FW transformation can be performed exactly. In this case, for solutions $\phi^{\pm}$ of Eq. (9), the corresponding projection by Eq. (18) is

$$\Phi^{\pm}=\frac{1}{2}\begin{bmatrix}\left(1+\dfrac{i}{2k_0}\dfrac{\partial}{\partial x}\log\dfrac{\mathscr{H}}{k_0}\right)\pm\dfrac{\mathscr{H}}{k_0}\\ \left(1-\dfrac{i}{2k_0}\dfrac{\partial}{\partial x}\log\dfrac{\mathscr{H}}{k_0}\right)\mp\dfrac{\mathscr{H}}{k_0}\end{bmatrix}\phi^{\pm}. \tag{26}$$

One can then define $(1\pm(i/2k_0)\partial_x\log(\mathscr{H}/k_0))\equiv e^{\beta\pm i\varphi}$, and choose for a first FW rotation $U^1=e^{-i\varphi\eta/2}$, to replace $\Phi^{\pm}$ with

$$U^1\Phi^{\pm}=\frac{e^{\beta/2}}{2}\begin{bmatrix}\sqrt{\dfrac{k_0}{\mathscr{H}}}\,e^{(\beta+i\varphi)/2}\pm\sqrt{\dfrac{\mathscr{H}}{k_0}}\,e^{-(\beta+i\varphi)/2}\\ \sqrt{\dfrac{k_0}{\mathscr{H}}}\,e^{(\beta-i\varphi)/2}\mp\sqrt{\dfrac{\mathscr{H}}{k_0}}\,e^{-(\beta-i\varphi)/2}\end{bmatrix}$$
$$\times(e^{(1/2)\log(\mathscr{H}/k_0)}\phi^{\pm}). \tag{27}$$

Two more rotations may then be performed, $U^2=e^{i\sigma\rho/2}$, where σ is chosen so that $\tanh(\sigma/2)=-\tan(\varphi/2)$, and $U^3=e^{\gamma\eta\rho/2}$, with $\gamma=[\log(\mathscr{H}/k_0)-\beta]$. Somewhat laborious, but straightforward algebra then produces, from Eq. (27), the result

$$U^3U^2U^1\Phi^+=\begin{bmatrix}1\\0\end{bmatrix}(e^{(1/2)\log(\mathscr{H}/k_0)}\phi^+)$$
$$U^3U^2U^1\Phi^-=\begin{bmatrix}0\\1\end{bmatrix}(e^{(1/2)\log(\mathscr{H}/k_0)}\phi^-)\ . \tag{28}$$

Because in this system the transverse operators all commute,

$$-i\,\frac{\partial}{\partial x}\,(e^{(1/2)\log(\mathscr{H}/k_0)}\phi^\pm)=\pm\mathscr{H}(e^{(1/2)\log(\mathscr{H}/k_0)}\phi^\pm), \tag{29}$$

so Eq. (28) is the desired result, relating the full solution to two components which evolve according to a diagonal, Hermitian one-way equation. The scalar evolution operator, recall, is the one given to leading order in Eq. (17).

Again, the answer is different from the range-dependent one-way operator of Ref. 3, but there appears to be a problem. Now the leading nontrivial correction looks like $+\alpha^2/8k_0^2$! The problem lies not with any ambiguity in the FW procedure, but with the perturbation series obtained in Eq. (25).

The assumption in constructing this series was that the operator X is small, and that additional range derivatives produce additional orders of the same small parameter. For the exponential potential in this example, that is not the case, and in fact $\dot{X}$ is of the same order as X. Thus the off-diagonal terms in B^{II} are really no smaller than the leading modification that appears in the diagonal term A^{II} in Eq. (25).

The resolution is to use, rather than the previous second rotation U^{II} a modified rotation

$$U^{\mathrm{II}}_{\mathrm{mod}}=\exp\left[-i\frac{\dot{X}}{2}(1-3X)\left(\frac{\eta+\rho}{2}\right)-\left(X^2-\frac{\dot{X}^2}{8}\right)\left(\frac{\eta\rho}{2}\right)\right].$$

This produces, instead of Eq. (25), the form

$$A^{\mathrm{II}}_{\mathrm{mod}}=\left\{1+X-\frac{X^2}{2}+\frac{\dot{X}^2}{8}+\frac{X^3}{2}+\frac{3X\ddot{X}}{4}+\frac{7X\dot{X}^2}{8}+\frac{1}{4}(2\dot{X}^2-\ddot{X})+\cdots\right\},$$
$$B^{\mathrm{II}}_{\mathrm{mod}}=\left\{\frac{4X^3}{3}+\frac{3X\ddot{X}}{4}+\frac{3X\dot{X}^2}{4}+\frac{1}{4}(2\dot{X}^2-\ddot{X})+\cdots\right\}, \tag{30}$$
$$C^{\mathrm{II}}_{\mathrm{mod}}=\left\{-\frac{3X^2\dot{X}}{2}-\frac{\dot{X}\ddot{X}}{8}+\frac{\dot{X}^3}{32}+\cdots\right\}.$$

Now, making use of the fact that $(2\dot{X}^2-\ddot{X})$ in this potential is a term of order $(\alpha/k_0)^3$, the off-diagonal terms are indeed one order higher than the leading correction to the diagonal term, so that term is robust against all truly higher perturbative corrections.

The modified rotation leading to Eqs. (30) is not as hopelessly obscure as it looks. The general forms of the exact rotations leading Eqs. (28) really depend, at low orders, only on the $\partial_x \log\mathscr{H}$ term, which has much the same form as in the standard one-way equation. Therefore, even without exact solutions, the latter form could have been used to infer the correct rotation, which is then expressed perturbatively to the desired order. This does, however, emphasize the caution that the off-diagonal terms need to be considered, and any given perturbation theory checked for consistency with the sound speed profiles on which it is to be used. Also, the second small-parameter classification of terms assumed here is physically more natural than the first. The perturbative tool of using a constant part of X equal to $(n^2-1)/2$ as the starting expansion parameter is awkward, relative to subsequent applications of the range derivative. The latter act multiplicatively on X, whereas the former is an additive relation. Therefore, for a general slowly varying n, if the derivative $\partial n/\partial x\sim\alpha$ at a point where $n=1$, then X will also be of order α in a neighborhood of that point. This is equivalent to the previous physical observation that, in a region of varying n, one could always make a local, boundary condition respecting transformation to coordinates in which the solution looks like a Hankel function. Therefore its leading deviation from the flat solution should resemble that implied by Eqs. (15) and (17).

The algebra presented here is much simpler than that of the general FW expansion,[11] because of the transverse symmetry. Physically, though, it is representative of the more general case. The examples in this section show how a difference of coordinates replaces the pure derivative expansion of the usual one-way equation (for range symmetry) with a nontrivial combination of evolution corrections and endpoint rescalings (for transverse symmetry). Conversely, for general coordinates in which the sound speed is not transversely symmetric, the sequences of FW rotations above generate, in addition to the terms shown, an operator series from commutators involving the transverse derivatives. These must then also be taken into account to obtain a diagonalization where reflections are removed at the desired, total perturbative order.

Finally, it is possible to see the remarkable result of virtual backscatter, which exists in the exact solution and is reproduced perturbatively. Thinking locally, one would expect the longitudinal and transverse wave numbers to be related to the frequency and the acoustic index of refraction as

$$k_x^2+k_y^2=k_0^2n^2.$$

In fact, for example when $\phi^\pm$ is given by Eq. (9) with $\nu=1/2$, the phase advances in *range* as the real part of the evolution operator in Eq. (15), which is exactly $\pm k_0 n$. Thus, the correct solution appears to possess $k_x^2=k_0^2n^2$, even though $k_y^2\neq 0$. This interpretation is exactly consistent with the one-way evolution of the components produced in the FW procedure as seen in Eqs. (17) and (29). One could place pressure release boundaries at $y=0$ and $y=-\pi/\alpha$ in this system and obtain a *nondispersive* mode in a waveguide.

The same considerations show what kind of errors can result if virtual backscatter is not taken into account. For instance, one could consider making a signal entirely from modes with $\nu=1/2$, with the timeseries

$$s(x_0,y,t)=\int_0^\infty d\omega\left(\frac{\omega}{\omega_0}\,e^{-\omega^2/2\omega_0^2}\right)$$
$$\times e^{-(\alpha x_0/2)+i(\omega/c_0\alpha)e^{\alpha x_0}}e^{-i\omega t}\,\sin\left(\frac{\alpha y}{2}\right) \quad (31)$$

on some starting range slice at x_0. For large $\omega_0/c_0\alpha$, this is the derivative of a Gaussian pulse, with center frequency ω_0, which may arbitrarily well approximate the derivative of a δ function. In the exponential sound speed profile, such a wave is perfectly nondispersive in the time domain. Though it has neither a well-defined group speed nor a well-defined phase speed, it is clear from Eq. (31) that

$$s(x,y,t)=e^{-(\alpha/2)(x-x_0)}s(x_0,y,t-t_0(x)),$$
$$t_0(x)=\frac{1}{c_0\alpha}\,(e^{\alpha x}-e^{\alpha x_0}). \quad (32)$$

Hence a recognizable pulse with fixed time dependence, up to an overall scale factor, travels with exactly the expected local speed $1/(dt_0(x)/dx)=c_0/n(x)=c(x)$. This behavior can be compared with what would be obtained by implementing the standard one-way equation of Ref. 3, without incorporating the effects of virtual backscatter as corrections to the sound speed profile. It will be assumed that individual frequency components are propagated independently, with energy conservation imposed at each step, and that the resulting solutions are summed to produce the final downrange pulse.

V. ENERGY CONSERVATION AND PHASES

The essential point is that many different solutions, having different frequency dependencies, are consistent with energy conservation. In the current example, this may be seen as follows. At any single frequency on any given range slice, the phase of a scalar field can be advanced with the local operator

$$\mathscr{H}_{\text{local}}(n)=k_0\sqrt{n^2+\frac{\partial_y^2}{k_0^2}}=k_0\sqrt{n^2-\frac{\alpha^2}{4k_0^2}}\,. \quad (33)$$

This evolution is always consistent with *some* energy-conserving solution. Specifically, it could have been obtained by evolving with the complete range-dependent operator from the FW procedure, in an artificial sound speed profile

$$n_{\text{art}}^2=n^2-\frac{\alpha^2}{4k_0^2}, \quad (34)$$

defined so that

$$\mathscr{H}_{\text{complete}}(n_{\text{art}})=k_0\sqrt{n_{\text{art}}^2+\frac{\alpha^2}{4k_0^2}+\frac{\partial_y^2}{k_0^2}}=\mathscr{H}_{\text{local}}(n). \quad (35)$$

Imposing energy conservation on this solution essentially amounts to inverting the FW transformations of Eqs. (26)–(28). This is true, independent of how the constraint is actually imposed, because by construction the pseudo-Unitary transformations are those that preserve downrange energy transport, and the projection defining Φ was chosen so that the full solution could be recovered from it. The resulting wave is no longer a simple Hankel function, but to leading changed order in α/k_0 it can be approximated by the function

$$\phi_{\text{art}}\sim e^{-\alpha x/2+i(k_0/\alpha)\sqrt{e^{2\alpha x}+\alpha^2/4k_0^2}}\,\sin\left(\frac{\alpha y}{2}\right), \quad (36)$$

which manifestly obeys the phase relation implied by Eq. (33) and conserves energy at this order. A set of such waves, producing the time series (31), can thus be inserted at large negative x, propagated downrange according to Eq. (35), and summed to produce a time series at large positive x. If the Fourier transform is evaluated in a stationary phase approximation to obtain some quantitative estimate of the dispersion, the time series is given, apart from overall normalizations and phase terms, by

$$s(x,y,t)\propto\frac{e^{-i\omega_0 t}}{(1-i\alpha/8k_0n(x))^{1/2}}$$
$$\times e^{-(\omega_0^2/4(1-i\alpha/8k_0n(x)))(t-t_x)^2}\,\sin\left(\frac{\alpha y}{2}\right). \quad (37)$$

The pulse width, as inferred from this representation, increases with range roughly as

$$\langle t-t_x\rangle\propto\frac{(1+(\alpha/8k_0n(x))^2)^{1/2}}{\omega_0}. \quad (38)$$

It is important that nothing fundamental about this result makes particular use of the exponential nature of the sound speed profile, or its large-range limits. The incorrectly predicted dispersion comes from improper frequency dependence of the *local operator which advances the phase*. Energy is conserved at each step, but the phase correlations among different frequencies in the full solution are corrupted if the frequencies are propagated according to Eq. (33). This example is merely one in which nondispersion of the exact solution makes the discrepency particularly evident. Further, while α/k_0 was assumed small to compare the results with perturbation theory, these exact solutions exist for any k_0, so range gradients could easily have been chosen for which this error is as large as desired. Thus it is only the relation of the sound speed profile and its range derivatives to the wavelength at any place which determines of what sign, and how large, the corrections from virtual backscattering will be.

VI. BOOTSTRAP BENCHMARK PROBLEMS

If evolution between two endpoints with nonuniform sound speed profiles is attempted, the form of the final FW projection defined by Eqs. (26)–(28), with the insertion of Eq. (15), gives an oversimple indication of how this projection must prepare the input, and normalize the output wave fields, to represent all refractive effects correctly. The remarkable thing is that, in a sense, this is all artificial. If the same wave field can begin and end in a region with uniform sound speed profile, all derivative terms in the endpoint projections vanish, and internal refraction is entirely handled by the diagonal evolution operator, leaving a single, constant normalization at the endpoints. The well-known benchmarks

with sound speed profiles appropriate to this discussion[2,8] were not chosen to provide this feature, but for evaluation of the FW procedure, it is highly desirable.

This section demonstrates a class of conformal transformations that can be varied to produce a range of models, while retaining simple endpoint behavior. They are chosen to reasonably resemble systems of physical interest, such as waveguides with different possible combinations of pressure release or rigid boundary conditions, sound speed profiles resembling fronts or local deviations, or similar characteristics. The same simple asymptotes that remove profile dependence from the endpoint rotations make it possible to use standard PE wave field starters, and will be shown even to make the problem of approximating the FW corrections, to a large extent, independent of how the rest of the (standard) one-way equation is implemented.

The conformal mappings that provide these features are a direct extension of the exact solution already considered. Start with a rectangular strip of the complex plane, bounded above by the real axis, which will represent the ocean surface. That plane can be mapped to a wedge by an exponential, with the real axis mapped to its positive half, as in the previous cases. Now, any complex polynomial which deforms that wedge but preserves the real axis can be followed by a logarithm, to map wedges back to strips. A specific example of such a transformation is

$$w = e^{\alpha z}, \qquad w' = w\,\frac{1+(w/r_-)^{p+s}}{1+(w/r_+)^{p}}, \quad z' = \frac{1}{\alpha}\log w'. \tag{39}$$

Here p, s, r_+, r_- can be any real numbers, as long as the resulting map is single valued in the region of interest. The transformation of the wedge (from w to w'), looks like the identity map near the origin (corresponding to the asymptotic left in the strip), and like a simple power law with power $1+s$ for $|w|$ large. The upper surface of the waveguide remains flat in the image, preserving both the geometry and the normal boundary condition of the ocean surface.

Rather than start with an isospeed system, as in Eqs. (4) and (5), it is convenient to choose an index n^2 in the starting coordinates $n^2(z,\overline{z})=|dz'/dz|^2$. The scale factor s sets the overall dilation of the coordinates in the far right from their original values. Since the coordinates on the asymptotic left remain unchanged, this provides the ratio of the speeds of sound in the two asymptotes of the starting system as

$$c_{x\to\infty} = \frac{c_{x\to-\infty}}{(1+s)}. \tag{40}$$

Similarly, r_- and r_+ determine, respectively, the onset positions of decrease or increase of the starting sound speed. The boundaries of coordinates z, $\overline{z}$ may be chosen simple and planar, and the system in the image coordinates z', $\overline{z'}$ will have constant sound speed but some nontrivial bathymetry determined by the boundary images under the mapping. More complex, piecewise-linear boundaries than these have already been handled with PE methods.[9]

A choice $s=0$ but $r_-\neq r_+$ represents a local deviation, between regions which are asymptotically the same. r_- and

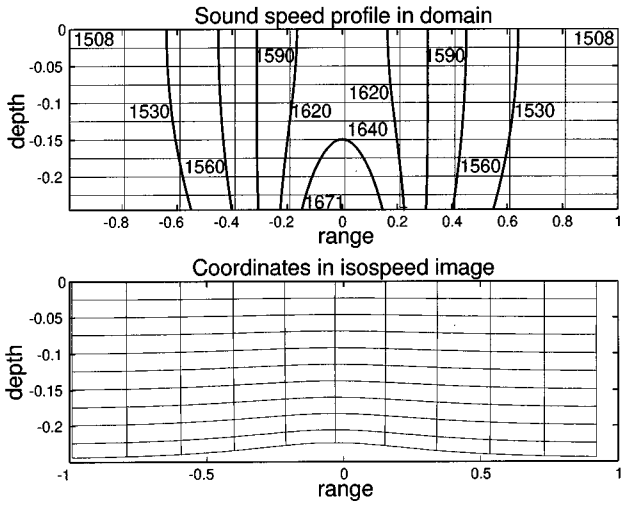


FIG. 1. Model of a local sound speed increase between two similar asymptotes: $r_-=1.01$, $r_+=0.99$, $s=0$, and $p=4$. Depth and range coordinates are plotted in arbitrary but identical units, in both frames. Lower frames in this and the following figures are isospeed at 1500 m/s, coordinate lines are images of the upper frame grid under the mapping.

r_+ can be chosen to make the deviation either fast or slow. An example with a faster central sound speed is shown in Fig. 1. The range and depth coordinates are plotted on the same scale, in arbitrary units. Large values of depth have been included, to make the corresponding bathymetry gradients visible in the mapped coordinates. The resulting distortion of the isovelocity contours from vertical could be nearly eliminated, if desired, by placing the lower boundary at a smaller aspect ratio of depth to range. [Sound-speed values in the upper corners of the upper frame do not satisfy Eq. (40), because the asymptotic regions have not been reached within the coordinate domain shown.] The lower frame is isospeed at 1500 m/s; coordinate lines are the images of the grid shown in the upper frame, under the conformal mapping. A choice $r_-=r_+$, $s\neq 0$, as in Fig. 2, crudely resembles a front, with monotonic change between asymptotically dif-

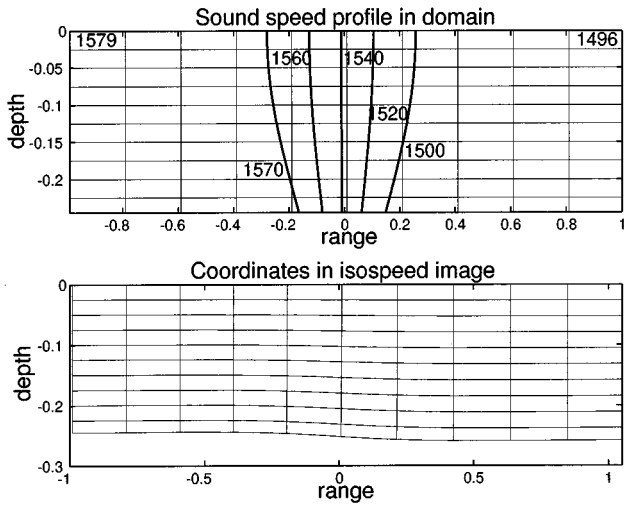


FIG. 2. A crude model of a front: $r_-=1.00$, $r_+=1.00$, $s=0.05$, and $p=4$.

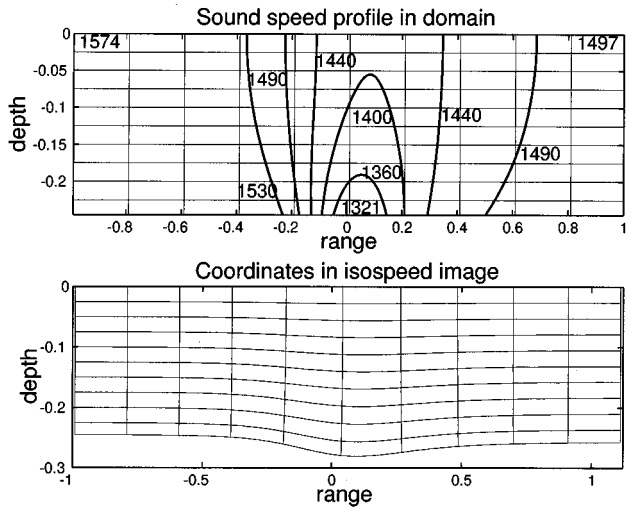


FIG. 3. A mixed model showing a local deviation between two dissimilar asymptotes: $r_- = 0.995$, $r_+ = 1.005$, $s = 0.05$, and $p = 6$. The sharpness and depth dependence of the transition have also been changed from the previous cases by choosing a different value of p.

ferent left and right values of the sound speed.

Varying the power p in any of these cases tunes the shape of the transition horizontally or vertically, thus changing the relation between the sound speed and its derivatives from point to point. As an example, a mixed model is shown in Fig. 3, incorporating a local deviation imposed on a monotonic change, and also using a different power p.

An alternative class of cases to consider is one with constant sound speed, as well as constant bathymetry, in the *the starting coordinates*, as in Eqs. (4) and (5), because the resulting modes can be solved exactly and are known to preserve energy exactly. The resulting image in these cases is not well-suited to separating refractive and boundary effects, because the region of changing bathymetry will also be the region of changing sound speed. However, the ability to compare the result to an exactly known solution makes it an eligible test of a numerical simulation which is expected to handle both.

VII. DISCUSSION

The computations of the preceding sections lead to a number of observations.

There is a logical problem with all PEs that approximate the usual one-way equation of Ref. 3 for range-dependent systems, which in some sense invalidates their formulation at any well-defined order of numerical approximation. Unless range derivatives are known to be smaller than all orders of transverse derivatives that are kept in expanding the one-way evolution operator, the resulting model is not a regulated small-parameter expansion in physical terms. The first observation is that FW fixes that, and in the case that the endpoint rotations are constant, the diagonal evolution operator may be constructed as a regulated expansion in gradients of the acoustic index of refraction.

On the other hand, in realistic ocean models, range derivatives of the sound speed profile are often not as large as depth derivatives, so properly, different small-parameter expansions should be used to describe the two classes of terms. Further, the perturbative FW diagonalization produces a power-series expression for the one-way operator which, in approximating the usual square root, is known not to be as useful numerically as rational function approximations.[5] This is why it is so important that FW corrects the *idealized* one-way equation, as can be seen from the fact that, in the range-independent case, it reduces to the optimal splitting-matrix decomposition of Ref. 15.

The FW corrections to the standard square root, as an operator series in transverse derivatives, necessarily have only coefficients containing range derivatives of the sound speed profile. Therefore, one may extract the leading-order terms with range derivatives as power-series corrections, and approximate the idealized equation (represented by all remaining terms not involving range derivatives) with rational functions or other means. In that way, an algebraically manageable expansion to low orders in range derivatives may be used consistently to correct a rational function expansion of depth derivatives that represents higher orders of a less-small parameter.

It was observed in Sec. I that, under evolution along sound speed gradients, the perturbation series of Ref. 11 has the same effect on phases as making many local, boundary condition respecting conformal transformations, and then supplying the appropriate Hankel functions in each little region to respect the angles of that region's resulting boundaries. Viewed from this perspective, a range-local evolution operator of the form (33) has no obvious relevance, because it does not even agree with the Hermitian part of the evolution implied by Eqs. (15) and (17). Therefore it should not obviously be a starting point to define the most general, physically correct PE.

However, while the operator $\mathscr{H}_{\text{local}}(n)$ may not be *fundamental*, it is *simple*. Further, for sound nearing infinite frequency, all terms in the expansion of any such $\mathscr{H}$ involving transverse or range derivatives of the sound speed profile (scaled by inverse powers of k_0 on dimensional grounds), would be suppressed out of existence. Therefore, in this limit, $\mathscr{H}_{\text{local}}(n)$ contains the only terms that are left. In that sense, the refraction of geometric acoustics represents the lowest-order nonlocality sampled by a sound wave.

The content of Sec. V, restated, is that one can retain the *form* of $\mathscr{H}_{\text{local}}(n)$, and still evolve consistently at noninfinite frequency, by substituting for $n(x)$ a frequency-dependent *effective acoustic index*, defined to leading order in α/k_0 by

$$n_{\text{eff}}^2(x,\omega_0) = n^2(x) + \frac{\alpha^2}{4k_0^2} + \cdots = n^2(x) + \frac{1}{4k_0^2}\left(\frac{\partial \log n}{\partial x}\right)^2 + \cdots . \tag{41}$$

As the frequency goes to infinity, the correction terms disappear, and one recovers the local, or ''bare'' acoustic index. For this reason, $\mathscr{H}_{\text{local}}(n)$ is also called the ''bare'' evolution operator.

Physically, virtual backscatter differs from real backscatter in that the former is essentially a coherent interference effect. A sound wave of finite wavelength, rather than propagating according the the sound speed at a point, effectively samples a smeared region. Because the propagation remains coherent, geometric-acoustics notions like surfaces of stationary phase and their associated normals (rays) do not become meaningless at noninfinite frequency, but to describe them in terms of local sound speed values, while capturing the altered relation between range and transverse components of the wave number, the local index $n(\mathbf{x})$ must be replaced by suitable frequency-dependent average of its values around the point $\mathbf{x}$. In the transversely symmetric case, the continuation of the power series presented in Sec. IV provides the whole expansion. In the more general case, the properly coordinate-independent average contains terms from the transverse-derivative algebra as well. The form of this algebra was the subject presented explicitly in Ref. 11.

ACKNOWLEDGMENTS

I wish to thank D. Wurmser and G. Orris for useful discussions of their work. In particular, I am indebted to them for making me aware that material related to the problem of benchmarking would serve a current need. I am also grateful to E. Westwood for comments, and to F. Jensen and A. M. Gleeson for pointers to the literature.

[1] F. D. Tappert, ''The parabolic approximation method,'' in *Wave Propagation and Underwater Acoustics*, edited by J. B. Keller and J. S. Papadakis (Springer-Verlag, New York, 1977), pp. 224–287.

[2] F. B. Jensen and C. M. Ferla, ''Numerical solutions of range-dependent benchmark problems in ocean acoustics,'' J. Acoust. Soc. Am. **87**, 1499–1510 (1990).

[3] F. Tappert, J. L. Spiesberger, and L. Boden, ''New full-wave approximation for ocean acoustic travel time predictions,'' J. Acoust. Soc. Am. **97**, 2771–2782 (1995).

[4] D. J. Thomson and N. R. Chapman, ''A wide-angle split-step algorithm for the parabolic equation,'' J. Acoust. Soc. Am. **74**, 1848–1854 (1983).

[5] M. D. Collins, ''A split-step Padé solution for the parabolic equation method,'' J. Acoust. Soc. Am. **93**, 1736–1742 (1993); ''Generalization of the split-step Padé solution,'' *ibid.* **96**, 382–385 (1994).

[6] M. D. Collins and R. B. Evans, ''A two-way parabolic equation for acoustic backscattering in the ocean,'' J. Acoust. Soc. Am. **91**, 1357–1368 (1992).

[7] J. A. DeSanto, ''Connections between the solutions of the Helmholtz and parabolic equations for sound propagation,'' in *Oceanic Acoustic Modeling*, edited by W. Bachmann and R. B. Williams (SACLANT Undersea Research Centre, La Spezia, Italy, 1975), pp. 43-1–43-17.

[8] D. J. Thomson, G. H. Brooke, and J. A. DeSanto, ''Numerical implementation of a modal solution to a range-dependent benchmark problem,'' J. Acoust. Soc. Am. **87**, 1521–1526 (1990).

[9] M. D. Collins, R. A. Coury, and W. L. Siegmann, ''Beach acoustics,'' J. Acoust. Soc. Am. **97**, 2767–2782 (1995).

[10] R. B. Evans, ''A coupled mode solution for acoustic propagation in a waveguide with stepwise variations of a penetrable bottom,'' J. Acoust. Soc. Am. **74**, 188–195 (1983).

[11] D. Wurmser, G. Orris, and R. Dashen, ''A new parabolic equation for range-dependent sound speeds derived using the Foldy–Wouthuysen transformation,'' J. Acoust. Soc. Am. **96**, 3343A (1994).

[12] For an excellent review of this and related topics, see H. Feshbach and F. Villars, ''Elementary relativistic wave mechanics of spin 0 and spin 1/2 particles,'' Rev. Mod. Phys. **30**, 24–45 (1958); for the more general context, and its relation to relativistic wave mechanics and quantum field theory, see C. Itzykhson and J-B. Zuber, *Quantum Field Theory* (McGraw-Hill, NY, 1980), pp. 185–195; a historical introduction is also provided in J. P. Costella and B. H. J. McKellar, ''The Foldy–Wouthuysen transformation,'' Am. J. Phys. **63**, 1119–1122 (1995).

[13] In nonrelativistic mechanics, classical field theory *is* quantum particle mechanics. FW arose in the early attempts to incorporate relativity into this paradigm. The unsuitability of relativistic classical fields for interpretation as quantum probability amplitudes was what made quantum field theory necessary. Although FW can be applied to the (operator valued) quantum fields, quantum interpretation has no direct relevance to the method itself.

[14] J. Corones, ''Bremmer series that correct parabolic approximations,'' J. Math. Anal. Appl. **50**, 361–372 (1975).

[15] S. T. McDaniel, ''Parabolic approximations for underwater sound propagation,'' J. Acoust. Soc. Am. **58**, 1178–1185 (1975).

[16] M. B. Porter, F. B. Jensen, and C. M. Ferla, ''The problem of energy conservation in one way models,'' J. Acoust. Soc. Am. **89**, 1058–1067 (1991); M. D. Collins and E. K. Westwood, ''A higher-order energy-conserving parabolic equation for range-dependent ocean depth, sound speed and density,'' J. Acoust. Soc. Am. **89**, 1068–1075 (1991).

[17] See, for instance, ''The Heard Island papers,'' J. Acoust. Soc. Am. **96**, 2327–2484 (1994).

[18] Y. Y. Lu and J. R. McLaughlin, ''The Riccati method for the Helmholtz equation,'' J. Acoust. Soc. Am. **100**, 1432–1446 (1996).

[19] The depth-independent case is used as an example here because definition of the resulting square root introduces no subtleties involving ordering of transverse derivatives, which are tangential to the point under consideration.

[20] E. L. Murphy, ''Ray representation of diffraction effects in the split-beam sound field,'' J. Acoust. Soc. Am. **43**, 610–618 (1968); ''Modified ray theory for the two-turning-point problem,'' *ibid.* **47**, 899–908 (1970).

[21] M. J. Buckingham, ''Acoustic propagation in a wedge-shaped ocean with perfectly reflecting boundaries,'' in *Hybrid Formulation of Wave Propagation and Scattering*, edited by L. B. Felsen (Nijhoff, Dordrecht, The Netherlands, 1984), pp. 77–105.

[22] M. J. Buckingham and A. Tolstoy, ''An analytical solution for benchmark problem 1: The 'ideal' wedge,'' J. Acoust. Soc. Am. **87**, 1511–1513 (1994).

[23] R. Bellman and R. Kalaba, ''Functional equations, wave propagation, and invariant embedding,'' J. Math. Mech. **8**, 683–704 (1959).

[24] The positive- and negative-frequency Hankel functions have been denoted here as $H^{\pm}$ rather than H^1 and H^2, to correspond to the complex exponentials of which they are the natural equivalents.

[25] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, New York, 1987), 2nd ed., pp. 251–256.

[26] These are general forms. Because of the particular x dependence of the sound speed profile, the action of the range derivative can always be reproduced by an operator involving only n^2 and transverse gradients. Since an arbitrary operator may be written as a sum of Hermitian and anti-Hermitian parts, and transverse symmetry of the sound speed profile removes the possibility of terms linear in transverse derivatives, the two terms must respectively take the forms $1\times$ and $i\times$ operators involving only the transverse Laplacian. Relations between $\mathscr{N}$ and $\mathscr{H}$ follow from energy conservation, and approximate forms for $\mathscr{H}$ can be obtained from the asymptotic expansions for the Hankel functions to arbitrary order.

[27] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, San Diego, 1980), Corrected and Enlarged Edition, p. 962.

[28] The evolution operator used here, $-i\partial/\partial x$, is given the sign appropriate to a momentum, because the Helmholtz equation is an elliptic, and not a hyperbolic, wave equation. Therefore, while the corresponding spatial operator is labeled $\hat{H}$, to indicate that the range direction is a direction of evolution, it is properly a momentum operator and not a Hamiltonian. This is opposite to the convention used in Ref. 11.

[29] Readers not familiar with Lie groups and Lie algebras may consult H. Georgi, *Lie Algebras in Particle Physics*, Frontiers in Physics series (Benjamin/Cummings, Menlo Park, 1982).

Source depth and the spatial coherence of ambient noise in the ocean

Michael J. Buckingham[a] and Nicholas M. Carbone
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0213

(Received 4 April 1996; accepted for publication 31 July 1997)

An analytical model is developed for the vertical coherence of the ambient noise field generated by a plane of sources at a finite depth beneath the ocean surface. To clarify the effects of source depth on the noise field, the relatively simple case of a semi-infinite ocean with an isovelocity profile is considered. The expression derived for the coherence is exact; it depends on the source depth explicitly, and it includes the homogeneous and inhomogeneous components of the field. When the sources are shallow, that is, the source depth is much less than a wavelength and the source–image pairs act as dipoles, the coherence is an oscillatory function of frequency, consistent with an earlier theory of noise coherence in deep water. With deeper sources, the dipole description fails and the coherence function becomes approximately independent of frequency. This change of character suggests that the spatial structure of the noise field at depth in the ocean could be inverted to yield information on the acoustic properties of the bubble sources associated with breaking surface waves.
 [S0001-4966(97)04311-7]

PACS numbers: 43.30.Bp, 43.30.Nb, 43.30.Re, 43.30.Pc [JHM]

INTRODUCTION

Ambient noise in shallow water is strongly influenced by the proximity of the seabed, which modifies the spatial structure of the noise field in the vertical.[1,2] This effect has been exploited recently as the basis of a noise inversion procedure for quantitatively determining the phase speeds of the compressional and shear waves in the seabed.[3,4]

In addition to the bottom, an important factor affecting the spatial coherence of ambient noise in the ocean is the depth of the noise sources. Concentrating on wind-related sources, particularly bubble formation arising from breaking wave events,[5–7] it is likely that the sources are at depths between a few centimetres to a meter or so, depending on surface conditions. At frequencies such that the source depth is a small fraction of a wavelength, a source and its negative image in the surface act as a dipole[8–10] and exhibit the characteristic dipolar radiation pattern, showing a null in the horizontal; but at higher frequencies, where the source depth is a significant fraction of a wavelength or greater, the dipole description is no longer valid. The transition from dipole behavior to two monopoles of opposite sign introduces some interesting features into the vertical coherence structure of the ambient noise field.

The earliest theoretical model for the spatial coherence of surface-generated noise in the ocean was developed by Cron and Sherman.[11,12] They represented the ocean as a semi-infinite half-space with a uniform sound speed profile, and assumed that the noise sources were distributed at random points in a layer lying immediately beneath the surface. Their treatment is approximate in that it neglects the spatially inhomogeneous component in the noise field; and the depth of the source layer was taken to be so small that it simply cancels out of their expression for the noise coherence function. Similarly, in several of the more sophisticated noise models that were to follow, the source depth is absent from the theoretical description of the coherence.

The present article evolved out of our attempts to invert the vertical coherence of ambient noise in shallow water for the geoacoustic parameters of the bottom.[4] To perform reasonable inversions, especially at the higher frequencies (i.e., around 600 Hz and above with a hydrophone spacing of 1 m), we found that it is necessary to include explicitly the source depth in our noise model. Indeed, by so doing, the depth of the sources became one of our inversion parameters, along with those describing the geoacoustic properties of the bottom.

To illustrate the effects of source depth on the vertical spatial structure of the noise a relatively simple model is developed here, which is an extension of the original deep-water model of Cron and Sherman;[11,12] that is to say, to clarify the relationship between source depth and vertical coherence, the influence of the bottom has been removed from the problem entirely. The analysis leads to an exact, closed-form expression for the vertical coherence of the noise in which the source depth appears explicitly as a parameter. This result for the coherence function includes terms representing both the homogeneous and inhomogeneous components of the noise field, although the latter is entirely negligible at receiver depths greater than a wavelength. In the limit of zero source depth, the homogeneous term reduces to the form derived by Cron and Sherman,[11,12] as required. A finite source depth, on the other hand, introduces a significant departure from Cron and Sherman's result.

[a]Also affiliated to: Institute of Sound and Vibration Research The University, Southampton SO17 1BJ, England.

I. PROPAGATION IN A SEMI-INFINITE ISOTROPIC OCEAN

A fundamental element of any ambient noise model is the Green's function representing the propagation between a single source and a receiver. In the case of interest here, in which a monopole source radiates sound beneath a pressure-release sea surface, the Green's function can immediately be stated as the difference between two spherical spreading terms, each of the form e^{jkR}/R, where k is the acoustic wave number, $j=\sqrt{-1}$, and R is the distance between receiver and source in one case or receiver and image in the other. This formulation of the Green's function was Cron and Sherman's[11,12] starting point, but, as they pointed out, it leads to unmanageable integrals in the analysis of the noise and forces the introduction of an approximation that we wish to avoid.

An alternative expression for the Green's function, G, of a single source is obtained by solving the Helmholtz equation using a cylindrical coordinate system with the origin at the surface and the axis vertical, passing through the receiver. Depth is positive downward. The details of the analysis are given in the Appendix. The solution for G is in the form of an inversion integral over horizontal wave number, p:

$$G=-\frac{Q}{4\pi j}\int_0^\infty \frac{p}{\eta}\{e^{j\eta(z-z')}-e^{j\eta(z+z')}\}J_0(pr)dp, \quad z>z', \tag{1}$$

where r is horizontal range, z and z' are receiver and source depths, respectively, $J_0(\cdots)$ is the Bessel function of the first kind of order zero, Q is the source strength, and the vertical wave number is

$$\eta=\sqrt{k^2-p^2}, \quad \mathrm{Im}(\eta)>0. \tag{2}$$

The inequality in Eq. (2) ensures that the solution for G converges at infinite depth. In Eq. (1) it is apparent that G is a function of frequency, since the radical η depends on the wave number, $k=\omega/c$, where ω is angular frequency and c is sound speed in the medium. This frequency dependence arises because G is the Fourier transform (with respect to time) of the acoustic pulse arrival at the hydrophone from a single, near-surface, transient source at horizontal range r. Such a source could be, for example, a breaking surface wave, which, through the radial oscillations of the associated subsurface bubbles, generates a pulse of sound lasting several seconds.

It is implicit in Eq. (1) that the source strength, Q, is independent of frequency and hence that the source is an impulse occurring at time $t=0$. Such a source has a white spectrum, which is an unrealistic representation of the sound from breaking waves. For our purpose, however, this is of no consequence because the normalization used in forming the noise coherence function is such that the source spectrum cancels out. Indeed, the fact that the coherence function depends on the propagation conditions but is independent of the spectral shape of the sources makes it attractive as the basis of noise inversion techniques for characterizing the ocean environment. (If two or more types of noise process were present, for instance, wind sources and shipping, then the spectral shapes of the sources would influence the coherence function. This, however, is not the situation we are considering.)

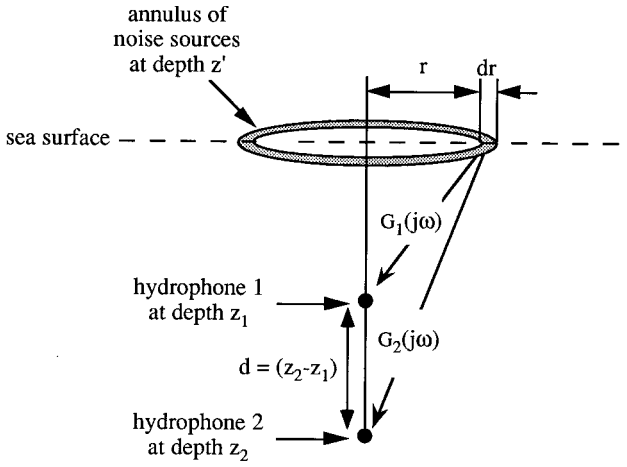


FIG. 1. Schematic showing a vertically separated pair of hydrophones centred on an annulus of near-surface noise sources. The Green's functions G_1 and G_2 are the Fourier transforms (with respect to time) of the velocity potential of a single noise pulse at each of the sensors, and hence are functions of angular frequency, ω.

Obviously, Eq. (1) can be integrated to yield the two familiar spherical spreading terms representing the source and its negative image in the surface. Although Eq. (1) is seemingly more complicated than these two simple terms, there is nevertheless an advantage to be gained by using the inversion integral expression for G when it comes to establishing a complete and exact model of the surface-generated noise field.

II. THE NOISE MODEL

As in previous models of surface-generated noise,[1,13] the sources are assumed to be independent and Poisson distributed in a plane beneath the sea surface. The surface is split into concentric annuli centred on the receiver pair, as shown in Fig. 1, and the contribution to the cross-spectral density of the noise at the receivers from the sources in each annulus is established from Carson's theorem. The cross-spectral density of the noise from the sources within all the annuli is then found by integrating over the whole surface.

This procedure, which is discussed in Ref. 13, leads to the following range integral for the cross-spectral density:

$$\overline{S_{12}}=4\nu\pi\int_0^\infty rG_1G_2^*\,dr, \tag{3}$$

where the overbar and the asterisk denote an ensemble average and complex conjugation, respectively, and ν is the mean rate of source pulses per unit area of surface. The subscripts 1 and 2 identify the Green's functions at receiver depths z_1 and z_2, respectively, from sources within the annulus at range r and of thickness dr. When the wave number formulation in Eq. (1) for the Green's functions is substituted into Eq. (3), the expression for the cross-spectral density takes the form of a triple integral:

$$\overline{S_{12}} = \frac{\nu Q^2}{4\pi} \int_0^\infty \int_0^\infty \int_0^\infty \frac{pp'}{|\eta|^2} F(p,z_1)F^*(p',z_2) \times J_0(pr)J_0(p'r)r\, dr\, dp\, dp', \quad (4a)$$

where p' is a dummy horizontal wave number, and

$$F(p,z) = e^{j\eta(z-z')} - e^{j\eta(z+z')} = -2je^{j\eta z} \sin \eta z'. \quad (4b)$$

The integral over range in Eq. (4a) is the Bessel function closure relation,

$$\int_0^\infty rJ_0(pr)J_0(p'r)dr = \frac{\delta(p-p')}{p}, \quad (5)$$

which is a convenient result, since the appearance of the Dirac delta function means that one of the two remaining integrals can be performed immediately, to yield

$$\overline{S_{12}} = \frac{\nu Q^2}{4\pi} \int_0^\infty \frac{p}{|\eta|^2} F(p,z_1)F^*(p,z_2)dp. \quad (6)$$

Equation (6) is an exact expression for the cross-spectral density that is valid provided both sensors lie below the noise sources. It can be further reduced by splitting the integration range into two regions, from zero to k and k to infinity. By making a change of variable, and with a little algebraic manipulation, the cross-spectral density then becomes

$$\overline{S_{12}} = \frac{\nu Q^2}{2\pi} \left\{ \int_0^k \frac{e^{-j\eta d}}{\eta} (1-\cos 2\eta z')d\eta + \int_0^\infty \frac{e^{-2\xi z_0}}{\xi} (\cosh 2\xi z' - 1)d\xi \right\}, \quad (7)$$

where

$$d = z_2 - z_1 \quad \text{and} \quad z_0 = (z_1+z_2)/2 \quad (8)$$

are the sensor separation and mean sensor depth, respectively. Since the first integral in Eq. (7) depends on sensor separation but not absolute position, it represents the homogeneous component of the noise field. The second integral is a function of the mean depth of the sensors and hence represents the spatially inhomogeneous noise. Notice that the source depth, z', appears explicitly in both integrals.

No approximations have been made in arriving at Eq. (7). If, now, the two terms containing the source depth are approximated to second order by their series expansions, the integrals can be expressed explicitly in terms of elementary functions. Partial integration then leads to the result

$$\overline{S_{12}} \approx \frac{\nu Q^2 k^2 z'^2}{\pi} \left\{ \frac{je^{-jkd}}{kd} + \frac{(e^{-jkd}-1)}{(kd)^2} + \frac{1}{(2kz_0)^2} \right\}; \quad (9)$$

and, to the same order of approximation, the power spectrum at each of the receivers is

$$\overline{S_{ii}} \approx \frac{\nu Q^2 k^2 z'^2}{2\pi} \left\{ 1 + \frac{1}{2(kz_i)^2} \right\}, \quad (10)$$

where the subscript $i=1$ or 2. Thus to this level of approximation, the coherence between the noise fluctuations at the two sensors is

$$\Gamma_{12} = \frac{\overline{S_{12}}}{\sqrt{\overline{S_{11}} \cdot \overline{S_{22}}}} \approx \frac{2\left\{ \frac{je^{-jkd}}{kd} + \frac{(e^{-jkd}-1)}{(kd)^2} + \frac{1}{(2kz_0)^2} \right\}}{\left\{ 1 + \frac{1}{2(kz_1)^2} \right\}^{1/2} \left\{ 1 + \frac{1}{2(kz_2)^2} \right\}^{1/2}}. \quad (11)$$

This expression for the coherence function is independent of the source depth, z', which cancels out when the cross-spectrum is normalized by the power spectra.

If the inhomogeneous terms are ignored, Eq. (11) becomes

$$\Gamma_{12} \approx 2\left\{ \frac{je^{-jkd}}{kd} + \frac{(e^{-jkd}-1)}{(kd)^2} \right\}, \quad (12)$$

which is Cron and Sherman's result[11,12] for the coherence of deep-water, surface-generated noise. The more complete expression, in Eq. (11), containing the inhomogeneous as well as the homogeneous terms, was derived by Isakovich and Kur'yanov[14] and, using a different approach, by Buckingham,[15] in both cases in connection with low-frequency ambient noise generation in the ocean. It is clear from Eqs. (9) to (11) that the inhomogeneous noise is negligible at sensor depths greater than a wavelength. At shallower depths, however, the inhomogeneous noise component may be a significant or even a dominant contributor to the field. For instance, in the case of infra-sonic noise at a frequency of 1 Hz and a depth of 40 m, the ratio of the inhomogeneous to homogeneous components in the power spectrum [Eq. (10)] is approximately 18:1.

The series approximation leading to Cron and Sherman's expression in Eq. (12) is valid when $2kz' < 1$, which is also the condition that identifies a source and its negative image in the surface as a dipole.[16] When this condition is not satisfied, that is, when the source depth is greater than a small fraction of a wavelength, Eq. (12) no longer provides an accurate description of the coherence. The more general expression in Eq. (7), on the other hand, represents the coherence exactly, however deep the source plane is beneath the surface.

III. EXACT TREATMENT OF THE SOURCE DEPTH

To investigate the effect of source depth on the noise coherence, the second-order series approximations for the trigonometric and hyperbolic functions in Eq. (7) must be abandoned. Returning to the exact result for the cross-spectral density of the noise in Eq. (7), the integrands may be expressed in terms of exponentials, yielding

$$\overline{S_{12}} = \frac{\nu Q^2}{4\pi} \left\{ \int_0^k [2e^{-j\eta d} - e^{-j\eta(d+2z')} - e^{-j\eta(d-2z')}] \times \frac{d\eta}{\eta} + \int_0^\infty [e^{-2\xi(z_0-z')} + e^{-2\xi(z_0+z')} - 2e^{-2\xi z_0}] \frac{d\xi}{\xi} \right\}. \quad (13)$$

Each of the terms in this expression is an exponential integral,[17] and each has a lower limit of zero, as a conse-

quence of which it diverges. However, all is not lost, for if the lower limit is replaced by ϵ and the limit taken as ϵ goes to zero, all the divergent terms cancel, leaving an expression for the cross-spectral density that is well-behaved.

To illustrate the cancellation, consider the exponential integral

$$\mathrm{Ei}(-\epsilon) = -\int_{\epsilon}^{\infty} \frac{e^{-x}}{x}\, dx, \tag{14a}$$

which in the limit can be written as

$$\lim_{\epsilon\to 0}\{\mathrm{Ei}(-\epsilon)\} = \gamma + \ln(\epsilon), \tag{14b}$$

where $\gamma = 0.577\,215\,7...$ is Euler's constant. Now, on taking the inhomogeneous term in Eq. (13), we find that

$$\begin{aligned} I &= \lim_{\epsilon\to 0}\int_{\epsilon}^{\infty}\{e^{-2\xi(z_0-z')} + e^{-2\xi(z_0+z')} - 2e^{-2\xi z_0}\}\,\frac{d\xi}{\xi} \\ &= \lim_{\epsilon\to 0}\{2\ln(2\epsilon z_0) - \ln[2\epsilon(z_0+z')] - \ln[2\epsilon(z_0-z')]\} \\ &= -\ln\left\{1 - \frac{z'^2}{z_0^2}\right\}, \end{aligned} \tag{15}$$

which is finite. Following a similar line of reasoning, the exponential integrals of imaginary argument in Eq. (13), representing the homogeneous field, can be expressed as

$$\begin{aligned} J &= \lim_{\epsilon\to 0}\int_{\epsilon}^{k}\{2e^{-j\eta d} - e^{-j\eta(d+2z')} - e^{-j\eta(d-2z')}\}\,\frac{d\eta}{\eta} \\ &= \ln\left\{1 - \frac{4z'^2}{d^2}\right\} + 2\,\mathrm{Ei}[-jkd] - \mathrm{Ei}[-jk(d+2z')] \\ &\quad - \mathrm{Ei}[-jk(d-2z')], \end{aligned} \tag{16}$$

a result which is also finite.

At this point, it is convenient to introduce the identity[18]

$$\begin{aligned} \mathrm{Ei}(-jx) &\equiv \int_{\infty}^{x}\frac{e^{-jx}}{x}\,dx \\ &= \gamma + \ln(x) - \mathrm{Cin}(x) - j\ \mathrm{Si}(x) + j\pi/2, \end{aligned} \tag{17}$$

where

$$\mathrm{Si}(x) = \int_0^x \frac{\sin x}{x}\,dx \tag{18}$$

is the sine integral and

$$\mathrm{Cin}(x) = \int_0^x \frac{(1-\cos x)}{x}\,dx. \tag{19}$$

With these expressions, in conjunction with Eqs. (15) and (16), the cross-spectral density in Eq. (13) becomes

$$\begin{aligned} \overline{S_{12}} = \frac{\nu Q^2}{4\pi}\Bigg(&\{\mathrm{Cin}[k(d+2z')] + \mathrm{Cin}[k(d-2z')] \\ &- 2\,\mathrm{Cin}[kd]\} - \ln\left[1 - \frac{z'^2}{z_0^2}\right] + j\{\mathrm{Si}[k(d+2z')] \\ &+ \mathrm{Si}[k(d-2z')] - 2\,\mathrm{Si}[kd]\}\Bigg). \end{aligned} \tag{20}$$

Equation (20) is our final formulation of the cross spectrum of the noise. Although it contains two special functions, $\mathrm{Cin}(\cdots)$ and $\mathrm{Si}(\cdots)$, these are entire functions (i.e., they show no singularities) that are particularly fast and accurate to compute using either a series expansion or a rational approximation,[18] depending on the value of the argument. Thus even though $\mathrm{Cin}(\cdots)$ and $\mathrm{Si}(\cdots)$ are defined as integrals, no numerical integration is necessary to evaluate any of the terms of Eq. (20). The power spectrum at either of the receivers is found from Eq. (20) by setting the element spacing, d, equal to zero:

$$\overline{S_{ii}} = \frac{\nu Q^2}{4\pi}\left\{2\,\mathrm{Cin}(2kz') - \ln\left[1 - \frac{z'^2}{z_i^2}\right]\right\}, \quad i = 1,2. \tag{21}$$

Note that, since no approximations are present in Eqs. (20) and (21), they are generally valid for any source depth. By performing a Taylor expansion to second order in the variable $2kz'$, these results reduce identically to the approximate forms in Eqs. (9) and (10). The truncated Taylor series are valid provided $2kz' < 1$, which is the condition that identifies a source and its negative image in the surface as a dipole. When this condition holds, Eqs. (9), (10), and the Cron and Sherman expression for the coherence in Eq. (12) are valid. However, for greater source depths or frequencies such that $2kz' > 1$, indicating nondipole behavior, the exact expressions in Eqs. (20) and (21) are the appropriate forms to use.

IV. EFFECTS OF SOURCE DEPTH ON COHERENCE

To form the coherence function, Γ_{12}, the cross-spectral density is normalized by the square root of the product of the power spectra, as defined in Eq. (11). Thus from Eqs. (20) and (21), the real and imaginary parts of the coherence function, including the inhomogeneous terms, are

$$\mathrm{Re}(\Gamma_{12}) = \frac{\{\mathrm{Cin}[k(d+2z')] + \mathrm{Cin}[k(d-2z')] - 2\,\mathrm{Cin}[kd]\} - \ln[1 - z'^2/z_0^2]}{\{2\,\mathrm{Cin}(2kz') - \ln[1 - z'^2/z_1^2]\}^{1/2}\{2\,\mathrm{Cin}(2kz') - \ln[11 - z'^2/z_2^2]\}} \tag{22a}$$

and

$$\mathrm{Im}(\Gamma_{12}) = \frac{\{\mathrm{Si}[k(d+2z')] + \mathrm{Si}[k(d-2z')] - 2\,\mathrm{Si}[kd]\}}{\{2\,\mathrm{Cin}(2kz') - \ln[1 - z'^2/z_1^2]\}^{1/2}\{2\,\mathrm{Cin}(2kz') - \ln[1 - z'^2/z_2^2]\}^{1/2}}. \tag{22b}$$

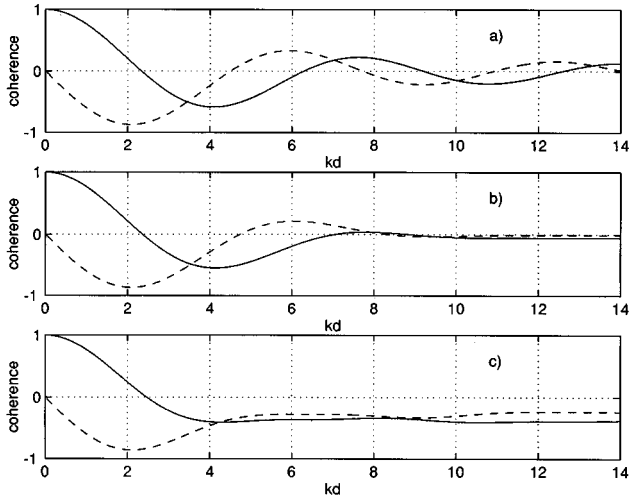


FIG. 2. Effect of increasing source depth on the coherence function. The solid and dashed lines are, respectively, the real and imaginary parts of Γ_{12}, as calculated from Eqs. (23). (a) $z'=0$ (Cron and Sherman); (b) $z'/d=0.25$; and (c) $z'/d=0.5$.

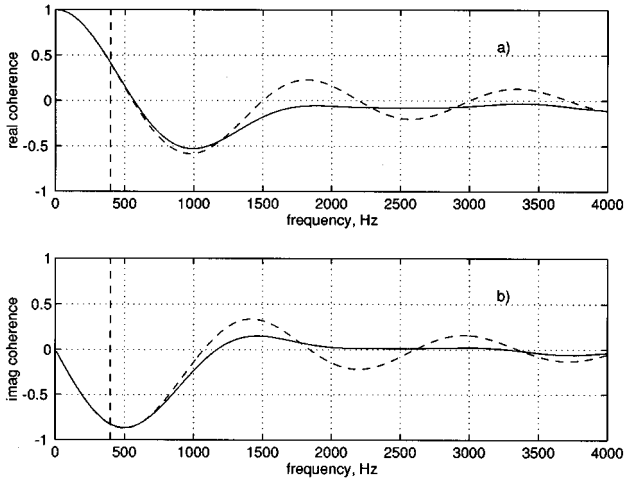


FIG. 3. (a) Real and (b) imaginary coherence curves as a function of frequency for an interelement spacing of $d=1$ m. The dashed curves are from Cron and Sherman's expression, Eq. (12), and the solid lines are from the exact result in Eqs. (23) for a source depth of $z'=0.3$ m. The differences between the curves in both panels occur to the right of the vertical dashed line, indicating the frequency at which $2kz'=1$. To the left of this vertical line, where the source-image pairs act as dipoles, the curves are indistinguishable.

For receiver depths where the inhomogeneous noise, represented by the logarithmic terms, is negligible, these expressions reduce to

$$\mathrm{Re}(\Gamma_{12})=\frac{\{\mathrm{Cin}[k(d+2z')]+\mathrm{Cin}[k(d-2z')]-2\,\mathrm{Cin}[kd]\}}{2\,\mathrm{Cin}(2kz')} \tag{23a}$$

and

$$\mathrm{Im}(\Gamma_{12})=\frac{\{\mathrm{Si}[k(d+2z')]+\mathrm{Si}[k(d-2z')]-2\,\mathrm{Si}[kd]\}}{2\,\mathrm{Cin}(2kz')}. \tag{23b}$$

The coherence in Eqs. (23) is a function of the two-dimensionless variables kd and z'/d. Three examples illustrating the effect of the source depth on the coherence are shown in Fig. 2. In the case considered by Cron and Sherman, where the source depth is infinitesimal and the source-image pairs act as dipoles, the real and imaginary parts of Γ_{12} are oscillatory in character, showing a sequence of zero crossings. As the source depth increases, the curves tend to flatten at the higher frequencies, where the oscillatory behavior is lost, and the higher zero crossings are absent. This behavior can be attributed to the failure, at the higher frequencies, of the dipole description of a source and its image in the sea surface. By way of contrast, at lower frequencies, the curves show the oscillations that are characteristic of surface dipoles, consistent with Cron and Sherman's model, which holds as an accurate representation of the coherence.

Further illustration of the effect of source depth is shown in Fig. 3, where the coherence is plotted as a function of frequency rather than the dimensionless quantity kd. For a source depth of just 0.3 m, it can be seen that the exact expression for the coherence deviates from Cron and Sherman's result above 400 Hz, which is the frequency where $2kz'=1$. Thus at frequencies below 400 Hz, where the surface-dipole description of a source and its image is appropriate, Cron and Sherman's expression for the coherence in Eq. (12) is adequate; but it fails at higher frequencies, where the source depth is comparable to or greater than the wavelength.

Breaking waves inject bubbles beneath the ocean surface. At the instant of formation the bubbles ring for a few milliseconds,[19] thus creating the major part of the wind-generated noise field. The depth of the acoustically active bubbles is uncertain, although an estimate of 1.5 m has been determined from inversions of wave-breaking sound[20] measured at wind speeds of 10 m/s or greater in the sea surface bubble layer.[21] If wave-driven sources do indeed penetrate to depths of order 1 m, then, judging by Figs. 2 and 3, the spatial properties of the ambient noise field could provide a useful measure of the source depth. Although speculative, this argument is supported by evidence that is present in several sets of noise coherence data collected from shallow water sites around the U.K.

Frequency-independent source depths were used to compute the curves in Figs. 2 and 3. It is possible, however, that the resonant bubble sources produced by wave breaking are distributed in depth according to their size, which scales inversely with the resonance frequency. If, as has been suggested by several authors,[22–24] a process of repeated bubble fracture is responsible for the bubble size distribution in the ocean, the bigger bubbles, generating the lower frequencies, may be formed at shallower depths than the smaller, higher-frequency bubbles. As an example of how such a mechanism might affect the coherence of the noise field, suppose that the source depth shows a frequency dependence of the form

$$z'=z'_0(1-e^{-f/f_0}), \tag{24}$$

where f_0 is the e-folding frequency. The function in Eq. (24) is shown in Fig. 4, where it can be seen that at low frequencies the source depth scales approximately linearly with frequency, but with increasing frequency approaches the limit-

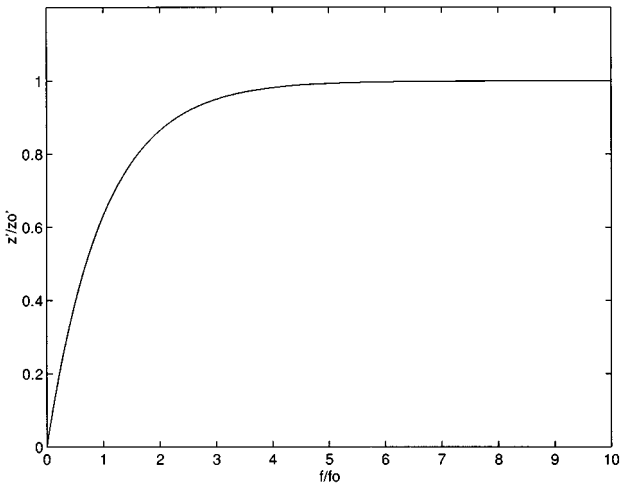


FIG. 4. Frequency dependence of the source depth, as given by the function in Eq. (24).

ing depth, z_0', asymptotically. Figure 5 shows three coherence functions computed from Eqs. (23), using the function in Eq. (24) for the source depth. The differences between the curves in Figs. 2 and 5 are sufficient to suggest that it may be possible to invert noise coherence data to establish the functional dependence of source depth on frequency.

V. CONCLUDING REMARKS

The analysis of deep-water, surface-generated ambient noise presented above illustrates that the depth of the sources is an important factor influencing the vertical coherence of the noise field. When the source depth is significantly less than a wavelength, a source and its negative image in the sea surface act as a dipole, and the real and imaginary parts of the coherence function oscillate about zero as the frequency increases. Such behavior is consistent with that originally predicted by Cron and Sherman[11,12] for the case of an infinitesimal source depth. For deeper sources, however, where the source depth is greater than a wavelength, the real and imaginary parts of the coherence are approximately independent of frequency, showing a form that is quite distinct from that of the Cron and Sherman curves.

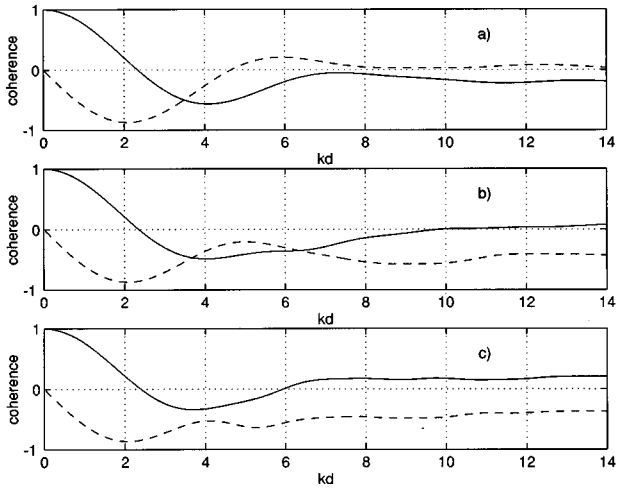


FIG. 5. Real (solid line) and imaginary (dashed line) coherence functions for three different values of the asymptotic source depth, z_0', computed from Eqs. (23) and (24). In all three panels the sensor separation is $d=1$ m and the e-folding frequency is $f_0=2$ kHz. (a) $z_0'=0.5$ m; (b) $z_0'=1.0$ m; and (c) $z_0'=1.5$ m.

Although the source-depth model is idealized, it is useful in providing some insight into the physics of subsurface noise sources and the spatial structure of the associated noise field at depth in the water column. Other factors, such as a rough sea surface, which have been neglected in the model, will also affect the spatial structure of the noise. Of course, in a full numerical treatment of the problem, all such mechanisms should be included.

An obvious feature that is absent from the model is the seabed, which in shallow water channels has a profound effect on the vertical structure of the noise field. The complicated boundary conditions associated with a realistic basement make it difficult, if not impossible, to develop closed-form analytical expressions for the noise coherence in shallow water that include the effect of a finite source depth. The natural alternative is an appropriate numerical technique. In fact, our shallow water noise inversions are based on an algorithm that includes the source depth as an inversion parameter.[25]

The effect of the source depth on the vertical coherence of shallow water noise is qualitatively similar to the behavior exhibited by the deep-water curves. With shallow sources, where the source–image pairs act as dipoles, the real and imaginary parts of the coherence are oscillatory functions of frequency,but both tend to become uniform with increasing frequency. The implication is that, in the higher-frequency regime, the spatial structure of the noise is governed by surface rather than bottom effects. This observation suggests that inversions of the noise coherence aimed at acquiring information about the shallow water environment could provide bottom typing at lower frequencies, where the sources act as dipoles, and source-depth characteristics at higher frequencies. In particular, the higher-frequency noise field may yield the functional relationship between the depth of the wind-driven sources (bubbles) and frequency. Such information is relevant to the use of passive acoustic techniques for inferring the void fraction profile in the sea surface bubble layer and the gas flux across the air–sea interface due to wave breaking.

ACKNOWLEDGMENTS

During the course of writing this paper, we benefited from many discussions on noise coherence with Dr. Grant Deane. This research was supported by the Ocean Acoustics Program, Office of Naval Research, under Grant No. N00014-91-J-1118, for which we are grateful.

APPENDIX: THE GREEN'S FUNCTION

The Green's function, G, of a point source at depth z' beneath a pressure-release surface is the solution of the Helmholtz equation:

$$\nabla^2 G + k^2 G = -Q\,\delta(\mathbf{r}-\mathbf{r}'), \tag{A1}$$

where k is wave number, $\mathbf{r}$ and $\mathbf{r}'$ are the position vectors of the receiver and source, respectively, and Q is the source strength. In cylindrical coordinates, with the axis vertical and passing through the source, this equation is

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right)+\frac{\partial^2 G}{\partial z^2}+k^2 G=-Q\frac{\delta(r)}{\pi r}\delta(z-z'), \quad \text{(A2)}$$

where r is horizontal range between source and receiver, and z is depth measured downward from the surface. Symmetry eliminates any azimuthal dependence.

To solve Eq. (A2), subject to the boundary condition

$$G=0 \quad \text{at} \quad z=0, \quad \text{(A3)}$$

we apply a Hankel transform of zero order in range, defined as

$$G_p=\int_0^\infty rG(r)J_0(pr)dr, \quad \text{(A4a)}$$

where p is the horizontal wave number and $J_0(\cdots)$ is the Bessel function of the first kind of order zero. The corresponding inverse transform is

$$G=\int_0^\infty pG_p J_0(pr)dp, \quad \text{(A4b)}$$

and the Hankel transform of the second derivative with respect to range is

$$\int_0^\infty \frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right)J_0(pr)dr=-p^2 G_p. \quad \text{(A4c)}$$

Thus the transformed version of Eq. (A2) is

$$\frac{\partial^2 G_p}{\partial z^2}+(k^2-p^2)G_p=-\frac{Q}{2\pi}\delta(z-z'). \quad \text{(A5)}$$

A Laplace transform over depth, z, is now applied to Eq. (A5), which yields

$$s^2 G_{ps}-\dot{G}_p(0)+(k^2-p^2)G_{ps}=-\frac{Q}{2\pi}e^{-sz'}, \quad \text{(A6)}$$

where s is the Laplace transform variable, the integration constant $\dot{G}_p(0)$ is the derivative of G_p normal to the sea surface at $z=0$, and we have used the convention of identifying a transform by subscripting with the transform variable. A second integration constant, $G_p(z=0)$ should also appear on the left of Eq. (A6), but has been set to zero by virtue of the pressure-release boundary condition in Eq. (A3). Equation (A6) provides an algebraic solution for the doubly transformed field G_{ps}:

$$G_{ps}=\frac{-Qe^{-sz'}+2\pi\dot{G}_p(0)}{2\pi(s^2+k^2-p^2)}. \quad \text{(A7)}$$

After applying a standard inverse Laplace transform to Eq. (A7), the unknown constant of integration is found to be

$$\dot{G}_p(0)=\frac{Q}{2\pi}e^{j\eta z'}, \quad \text{(A8)}$$

and the solution for the Hankel transformed field is

$$G_p=-\frac{Q}{2\pi\eta}\{u(z-z')\sin\eta(z-z') - u(z)e^{j\eta z'}\sin\eta z\}, \quad \text{(A9)}$$

where

$$\eta=\sqrt{k^2-p^2} \quad \text{(A10)}$$

is the vertical wave number and $u(..)$ is the unit step (Heaviside) function. On taking the inverse Hankel transform of Eq. (A9), we find that

$$G=-\frac{Q}{4\pi j}\int_0^\infty \frac{p}{\eta}\{e^{j\eta(z-z')}-e^{j\eta(z+z')}\}J_0(pr)dp, \quad z>z', \quad \text{(A11)}$$

which is the solution for the field in Eq. (1) in the text.

An alternative, and somewhat briefer, approach to deriving Eq. (A10) is simply to perform a Hankel transform of the familiar spherical spreading terms representing the source and its image in the surface. The inverse transform for G is then found to be exactly as shown above.

[1] M. J. Buckingham, ''A theoretical model of ambient noise in a low-loss, shallow water channel,'' J. Acoust. Soc. Am. **67**, 1186–1192 (1980).
[2] W. A. Kuperman and F. Ingenito, ''Spatial correlation of surface generated noise in a stratified ocean,'' J. Acoust. Soc. Am. **67**, 1988–1996 (1980).
[3] M. J. Buckingham and S. A. S. Jones, ''A new shallow-ocean technique for determining the critical angle of the seabed from the vertical directionality of the ambient noise in the water column,'' J. Acoust. Soc. Am. **81**, 938–946 (1987).
[4] M. J. Buckingham, G. B. Deane, and N. M. Carbone, ''Determination of elastic sea floor parameters from shallow-water ambient noise,'' in *Proceedings of the Second European Conference on Underwater Acoustics*, edited by L. Bjørnø (European Commision, Luxenbourg, 1994), pp. 19–25. Commission, Lyngby, Denmark, pp. 19–25.
[5] B. R. Kerman, *Sea Surface Sound: Natural Mechanisms of Surface Generated Noise in the Ocean* (Kluwer, Dordrecht, 1988), 639 pp.
[6] B. R. Kerman, *Natural Physical Sources of Underwater Sound: Sea Surface Sound (2)* (Kluwer, Dordrecht, 1993), 750 pp.
[7] M. J. Buckingham and J. R. Potter, *Sea Surface Sound '94: Proceedings of the III International Meeting on Natural Physical Processes Related to Sea Surface Sound* (World Scientific, Singapore, 1995), 494 pp.
[8] J. R. Urick, *Ambient Noise in the Sea* (Peninsula, Los Altos, CA, 1986).
[9] R. M. Kennedy, ''Sea surface dipole sound source dependence on wave-breaking variables,'' J. Acoust. Soc. Am. **91**, 1974–1982 (1992).
[10] L. Ding and D. M. Farmer, ''On the dipole acoustic source level of breaking waves,'' J. Acoust. Soc. Am. **96**, 3036–3044 (1994).
[11] B. F. Cron and C. H. Sherman, ''Spatial correlation functions for various noise models,'' J. Acoust. Soc. Am. **34**, 1732–1736 (1962).
[12] B. F. Cron and C. H. Sherman, ''Spatial correlation functions for various noise models,'' J. Acoust. Soc. Am. **38**, 885 (1965).
[13] M. J. Buckingham, ''On surface-generated ambient noise in an upward refracting ocean,'' Philos. Trans. R. Soc. London, Ser. A **346**, 321–352 (1994).
[14] M. A. Isakovich and B. F. Kur'yanov, ''Theory of low-frequency noise in the ocean,'' Sov. Phys. Acoust. **16**, 49–58 (1970).
[15] M. J. Buckingham, ''Infrasonic ambient noise in the ocean due to atmospheric pressure fluctuations on the surface,'' J. Acoust. Soc. Am. **88**, 984–994 (1990).
[16] P. M. Morse and K. U. Ingaard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), 927 pp.
[17] N. N. Lebedev, *Special Functions and their Applications* (Prentice-Hall, Englewood Cliffs, NJ, 1965), 308 pp.
[18] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), 1046 pp.
[19] M. S. Longuet-Higgins, ''Bubble noise mechanisms—A review,'' in *Natural Physical Sources of Underwater Sound*, edited by B. R. Kerman (Kluwer, Dordrecht, 1993), pp. 419–452.
[20] M. J. Buckingham, ''On acoustic transmission in ocean-surface

waveguides,'' Philos. Trans. R. Soc. London, Ser. A **335**, 513–555 (1991).
[21] D. M. Farmer and S. Vagle, ''Waveguide propagation of ambient sound in the ocean-surface bubble layer,'' J. Acoust. Soc. Am. **86**, 1897–1908 (1989).
[22] M. S. Longuet-Higgins, ''The crushing of air cavities in a liquid,'' Proc. R. Soc. London **439**, 611–626 (1992).
[23] W. K. Melville, ''The dynamics and acoustics of breaking waves,'' J. Acoust. Soc. Am. **91**, 2322(A) (1992).
[24] T. G. Leighton, M. F. Schneider, and P. R. White, ''Study of bubble fragmentation using optical and acoustic techniques,'' in *Sea Surface Sound '94: Proceedings of the III International Meeting on Natural Physical Processes Related to Sea Surface Sound*, edited by M. J. Buckingham and J. R. Potter (World Scientific, Singapore, 1995), pp. 414–428.
[25] N. M. Carbone, G. B. Deane, and M. J. Buckingham, ''The compressional and shear wave speeds of a seabed in shallow water determined from ambient noise measurements,'' in *Sea Surface Sound '94: Proceedings of the III International Meeting on Natural Physical Processes Related to Sea Surface Sound*, edited by M. J. Buckingham and J. R. Potter (World Scientific, Singapore, 1995), pp. 32–44.

Source track localization via multipath correlation matching

Evan K. Westwood and David P. Knobles
Applied Research Laboratories, The University of Texas at Austin, P.O. Box 8029, Austin, Texas 78713-8029

(Received 23 April 1997; accepted for publication 10 June 1997)

A method for determining the track of a source emitting continuous broadband acoustic energy in an oceanic waveguide is presented. The method involves cross correlating the measured signals at horizontally separated receivers over a period of time and identifying the traces on the resulting correlogram in terms of the ray paths at each receiver that produce them. Range and bearing information is contained in the structure of the multipath correlation traces. Environmental parameters are used as inputs to a ray model to obtain the ray travel times t as functions of range R. The $\overline{t}_j(R)$ functions for the ray multipaths are used to obtain simulated correlogram time delays given a source track and receiver geometry. Constant-velocity, constant-depth source tracks are parametrized by four variables, and a nonlinear optimization algorithm is used to find the track that provides the best fit between measured and simulated correlation traces. The method is applied to measurements of a passing surface ship in the shallow waters of the English Channel. Successful localization is obtained using 4-min intervals and correlation traces produced by rays with 3–15 traversals of the water column. Receiver separation, unknown because of experimental uncertainties, is also obtained. © *1997 Acoustical Society of America.* [S0001-4966(97)04911-4]

PACS numbers: 43.30.Wi, 43.30.Zk, 43.60.Gk [DLB]

INTRODUCTION

A variety of approaches for localizing sources of acoustic energy in the ocean have been developed. Each of the approaches tends to be oriented toward particular sensor geometries, source radiation assumptions, and/or set of propagation characteristics. Sensor geometries may be single elements, vertical arrays, or horizontal arrays; source radiation characteristics may be single frequency (cw), continuous broadband, or impulsive broadband; and the acoustic propagation may be best described in terms of normal modes or rays. In addition, some approaches assume a particular type of source motion, while others attempt to localize the source at each time instant independently. Considerable work has been done on the general problem of source localization given a set of time delay of arrival (TDOA) measurements (see Refs. 1–5). Matched-field processing constitutes another major approach to the problem (see Ref. 6 and references therein). A third general approach is space–time integration, where the plane-wave or matched-field beamformer outputs produced by candidate source tracks are matched to the measured data, and the optimal source track is determined via grid searches, simulated annealing, Newton's method, or a combination of these (see Refs. 7–9).

This article presents an approach to source track localization that uses the correlation between broadband signals that travel from a source to two or more receivers along multiple ray paths. Horizontal receiver separation provides bearing information, and the range dependence of the source–receiver eigenray arrival structure provides range information. The approach requires that the correlation traces observed on an imaged display of correlation coefficient versus time delay and time (a correlogram) be identified in terms of the ray paths at each receiver that produce them. The water depth, the measured sound-speed profile, and, in some cases, the seabed structure are used to obtain ray travel time $\overline{t}$ as a function of range R for each ray path j involved. For a candidate source track, the $\overline{t}_j(R)$ functions are used to compute simulated correlogram time delays. The localization method then uses a nonlinear optimization algorithm to find the constant-velocity, constant-depth source track that produces the best fit between simulated and measured correlation time delays.

The applicability of the method is limited mainly by the source–receiver propagation characteristics. The method may be used when the propagation can be described in terms of a set of eigenrays with time delay differences that are resolvable using the bandwidth available. In general, such a description applies when the ratio of the range to the water depth is not too large, say less than 20–40. In addition, when bottom-interacting ray paths are involved, the environment is assumed to be range independent. Modifications required to handle the range-dependent problem are discussed in Sec. IV. The assumption of straight-line source tracks is valid for most practical situations, especially when the overall track is pieced together from multiple localizations over short time periods. Another requirement for the method to be applicable is that the source must have a continuous broadband component; otherwise, multipath time delays cannot be determined.

Although many aspects of the present method are common to other methods, the originality of the present work lies in the combination of the following: (1) use of correlation traces produced by ray multipaths of arbitrary order; (2) use of the measured sound speed profile to trace rays, rather than assuming straight-line ray paths; (3) parametrization of the source track over each (possibly short) time period as constant velocity and constant depth; (4) exploitation of the broadband source spectrum for source localization in an efficient manner; and (5) successful application of the method

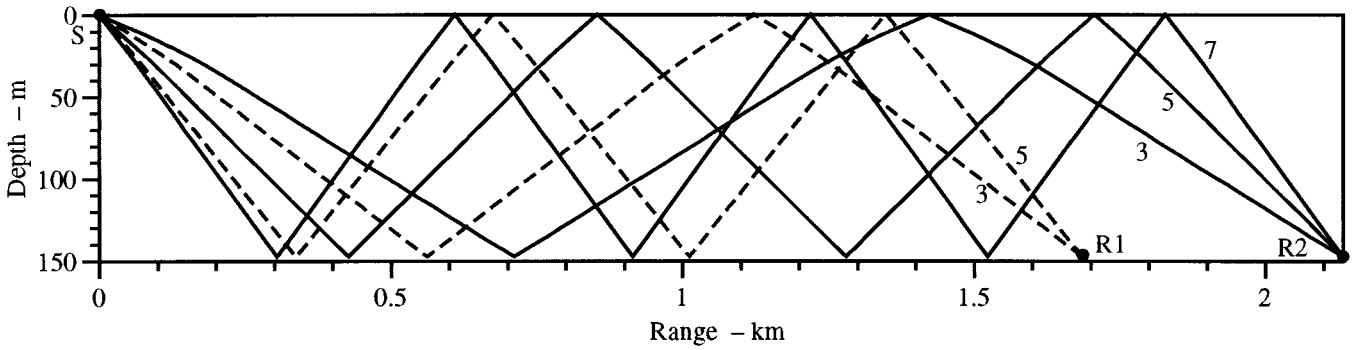


FIG. 1. Source–receiver eigenrays for two ranges in the downward-refracting, shallow-water environment described in Sec. II. Rays are labeled according to their number of water column traversals.

to data measured in an environment characterized by complex multipath propagation. The present method is an outgrowth of methods that use direct-path ray correlations to localize both sources and receiver elements.[10,11]

This article is organized as follows. In Sec. I we describe the source track localization method. In Sec. II we apply it to measured data from a shallow-water area in the English Channel. Section III contains an analysis of the effects of various types of mismatch on the accuracy of the method. Finally, Sec. IV discusses extensions of the method and summarizes our results.

I. SOURCE TRACK LOCALIZATION METHOD

The steps required by the localization method are (1) data processing, (2) environmental modeling, (3) correlogram trace identification and sampling, and (4) source track localization. Although we describe the method and present an application for the case where the number of receivers $N_R=2$, more receivers and correlograms [up to $N_R(N_R-1)/2$] may be used. For horizontally separated receivers, three or more noncollinear receivers are required to break the ambiguity in source localization inherent with a line array.

A. Data processing

Data processing consists of cross correlating the time series measured at pairs of receivers and imaging the normalized correlation coefficient as a function of time delay and recording time. In practice, the time series at each receiver $r(t)$ is digitized at a sample rate of f_s and fast Fourier transformed (FFTed), N_{FFT} points at a time, to obtain the received spectrum $R(f)$. The time interval of each FFT output is $T_w=N_{\mathrm{FFT}}/f_s$. The cross correlation $c(\tau)$ of the time series at receivers 1 and 2 is obtained as the inverse FFT of the product $C_{12}(f)=R_1(f)R_2^*(f)$, where the superscript * indicates complex conjugation. The time delay τ varies from $-T_w/2$ to $T_w/2$.

Spectral whitening of the received spectra is required when processing measured data because tonal components (''tonals'') are present in the spectra of typical acoustic sources. Tonals in the received spectra produce tonals in the correlation, which results in banding in the imaged correlogram. For spectral whitening, we divide each spectral component of C by its magnitude, leaving only the phase information. Normalization by the power in each signal is also applied in the frequency domain such that the correlation coefficient lies between -1 and 1.

B. Environmental modeling

The localization method requires computation of the ray travel time, which is denoted $\underline{t}$ to distinguish it from the recording time t, as a function of horizontal range R for each ray path j from the assumed source depth z_s to the receiver depth z_r. We use the range-invariant ray model GAMARAY[12,13] to trace approximately 100 rays over an interval of grazing angles θ and form tables of $(R,\underline{t})_j$. The sound-speed profile in the water column and the water depth are required inputs. Ray paths are characterized by their initial direction from the source (up or down) and their number of surface and bottom interactions. For our applications, where the receivers lie on the ocean bottom, the functions $R_j(\theta)$ and $\underline{t}_j(R)$ are monotonically increasing as long as the rays do not refract in the ocean bottom. The range R for each ray path varies from zero to a maximum ''cut-off'' value $R_{\max}$, which is finite when the profile is not isospeed.

For sources near the surface and receivers on the bottom, it is convenient to parametrize a ray path by the number of times it traverses the water column, which is an odd number. Figure 1 illustrates the dominant ray paths for two source–receiver ranges in the shallow-water environment that will be discussed in Sec. II. Each ray path is labeled according to its number of water column traversals. Direct-path rays do not arrive at the receivers because the ranges are greater than $R_{\max}$ for the direct path. The seven-traversal ray is stronger at the second receiver because its interaction angle at the bottom is shallower. Additional references to Fig. 1 will be made when we analyze measured data in Sec. II.

For ranges beyond one or two water depths and small ratios of z_s to water depth H, it is appropriate to merge the two paths that differ by a surface reflection by setting $z_s=0$ in the ray-tracing procedure. In our experience, the travel time difference between the direct and surface-reflected paths from near-surface sources causes multipath correlations to broaden somewhat, but the individual paths cannot be distinguished. Thus, we choose points in the middle of the

slightly broadened correlation traces and set $z_s = 0$, the average of the actual source depth and its image above the surface. For receivers on the bottom, the direct and bottom-reflecting multipaths also merge, resulting in only one ray path with a given number of ocean traversals.

The composition of the ocean bottom needs to be known only to the degree required to determine which ray paths dominate. A relatively hard bottom, consisting of sand, for example, causes specularly reflected rays to dominate; a relatively soft bottom, consisting of silt or silty clay, for example, causes bottom-penetrating rays to dominate. The dominant ray path, specularly reflecting or bottom penetrating and turning, is then used when tracing the rays for all bottom interactions (except the last, where the downward-arriving energy at the receiver is assumed to dominate any bottom interaction). Subsurface layering would require an estimation of the dominant reflective layer.

C. Correlogram trace identification and sampling

The step in the procedure most prone to error is identification of the ray paths that produce the traces observed on the measured correlogram(s). If the closest point of approach (CPA) of the source is within five or seven water depths, the direct path is typically present, and identification is fairly straightforward. However, when the CPA range is larger than $R_{1\,\max}$, the lowest order ray paths have three or more ocean traversals, and identification can be difficult. In Sec. II we point out several clues in the correlation trace structure that help clarify the identification, and in Sec. IV we point to extensions of the method in which different hypotheses regarding the path order would be tested.

Correlation traces are identified in the form $p_1 \times p_2$, where p_1 and p_2 identify the ray paths at receivers 1 and 2, respectively, in terms of the number of water column traversals (see Fig. 1). The cross-correlation convention is that the time delay $\tau = t_1 - t_2$, where t_i is the ray travel time to receiver i. Because rays with more traversals have larger travel times, the 1×3 correlation trace tends to appear at negative τ, while 3×1 tends to appear at positive τ.

Localization accuracy is enhanced by sampling traces corresponding to rays with different vertical angles. This fact can be understood by considering the contrary case: If all rays arrive horizontally, only bearing information exists, and an ambiguity exists between the CPA range and the source velocity. Therefore, as the range increases and the lowest order ray paths arrive at shallower angles, it is advantageous to sample as many of the higher order ray path correlations as possible, such as 3×3, 5×5, or 3×5.

In our implementation, different numbers of points can be taken from each correlation trace, which is especially convenient for time intervals during which certain correlation traces appear or disappear. The assumption of a straight-line source track makes it unnecessary to sample the traces simultaneously.

D. Source track determination

The source track over the time period covered by the sampled correlation traces is assumed to be at constant depth and velocity. The depth is assumed to be known, and for our applications we assume it is near the surface. The track is parametrized by four quantities: speed v, heading ϕ, and horizontal coordinates x_0 and y_0 at an arbitrary reference time t_0. When the relative receiver locations are not known, we add the (x,y) locations of the receivers to the list of source track parameters to be optimized.

The inputs to the localization algorithm are (1) a list of N (t, τ_m) points from any number of traces on the measured correlogram(s), (2) for each point, the ray paths that produce the correlation peak and the two receivers used in the correlation, (3) tables of $(t,R)_j$ for each ray path involved, sampled at approximately 100 points, and (4) initial values for the parameters to be optimized. For each candidate track, the source coordinates and source–receiver ranges at the sample times are computed from simple geometry. The $(t,R)_j$ tables are then fit with cubic splines, and the travel times t_1 and t_2 from the source to the two receivers are computed for each (t, τ_m) point. The ''simulated'' time delay is then simply $\tau_s = t_1 - t_2$. Finally, the mean-squared error between the measured and simulated correlation time delays,

$$e = \sum_{i=1}^{N} (\tau_m - \tau_s)^2, \tag{1}$$

is computed and used as a measure of localization accuracy. The nonlinear optimization algorithm LEASTSQ in the MATLAB[14] software package is then used to find the source track that minimizes e. Analytic derivatives of e with respect to the unknown parameters are provided to the algorithm so that less efficient numerical derivatives need not be computed.

The initial guess supplied to the optimization algorithm does not appear to be critical, although a good initial estimate obviously reduces the number of iterations required. A brief description of how the results depend on the initial guess for the experimental data case will be given in Sec. II C. An estimate of the heading ϕ can be obtained from the bearing of the source either at the time of CPA or at early and late times when the bearing is nearly constant. The bearings, in turn, may be obtained from the correlation time delays using the sound speed and receiver separation.

II. APPLICATION TO SHALLOW-WATER MEASUREMENTS

A. Experimental data

The measured data to which we apply the source track localization method were recorded in the English Channel during the Shallow Water Active Measurements Program (SWAMP) III experiment, conducted jointly by Naval Research Laboratory; Applied Research Laboratories, The University of Texas at Austin; and SACLANTCEN in July 1992. Acoustic data were measured at three hydrophones on the ocean bottom. The measured sound-speed profile is downward refracting and is plotted in Fig. 2(a). The upper 30 m of the ocean bottom consists of fine, hard-packed sand. Its estimated surficial sound speed, density, and attenuation are 1720 m/s, 1.95 g/cm^3, and 0.40 dB/m/kHz, respectively. Figure 2(b) shows the two receiver locations, which were esti-

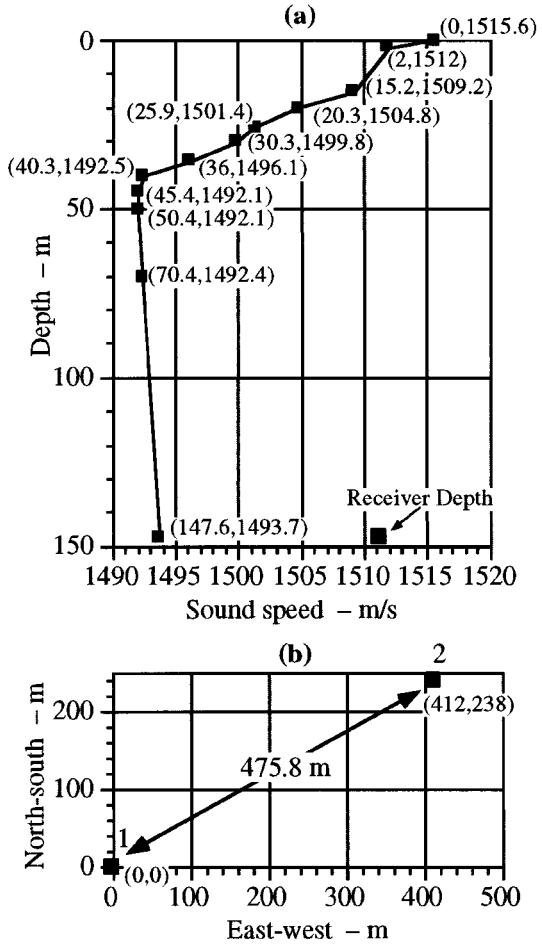


FIG. 2. (a) Sound-speed profile measured during the SWAMP III experiment in the English Channel. (b) Plan view of bottom-mounted receivers used in the source localization.

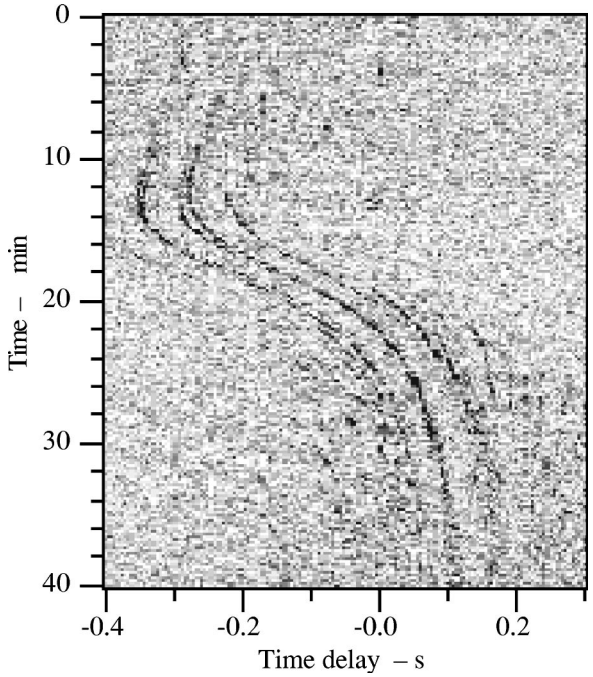


FIG. 3. Overall view of 40-min correlogram measured in the English Channel during the passage of a ship.

mated during their deployment. The receivers are separated by about 438 m, which corresponds to about three water depths. The sample rate was $f_s = 2000$ Hz, and the chosen FFT length was $N_{\mathrm{FFT}} = 4096$.

We examine in detail a 40-min period during which a ship passed near the receivers. To convey the overall structure of the correlogram, we first present in Fig. 3 a single correlogram that covers the entire time period. One of every five FFTs has been taken, and band-limited correlations (250–500 Hz) have been enveloped. The vertical and horizontal axes are recording time t and correlation time delay τ, respectively.

Many traces may be observed on the experimental correlogram in Fig. 3. In general, they evolve from negative to positive time delay τ, which indicates that the source was closer to receiver 1 at early times and closer to receiver 2 at later times. Near min 13 the traces reach their most negative point, which corresponds to the source being at endfire. At min 22 the strongest traces cross $\tau = 0$, which indicates that the source is at broadside.

To obtain more resolution, we re-process the data using every FFT and a 10- to 600-Hz frequency band, and we split the 40-min period into two sections. Detailed views of the two sections of the correlogram are shown in Fig. 4. The horizontal grid lines and the dots indicate the 4-min periods and the (t, τ) points that are used in the localization. The traces are labeled according to the ray paths that produce them (see Sec. II C and Fig. 1 for ray path nomenclature).

B. Multipath trace identification

Although the structure of the multipath correlations in Fig. 4 is complex, certain patterns exist that make the analysis more simple. In Fig. 4(a) five distinct groups of traces can be observed. The leftmost group consists of 3×7, 5×9, 7×11, and so forth, where the ray at receiver 2 has four additional ocean traversals (two additional bottom reflections) than does the ray at receiver 1. The next group to the right, 3×5, 5×7, 7×5, and so forth, are traces produced by rays with differences of two traversals or one bottom reflection. The middle group is made up of correlations between rays with the same paths: 3×3, 5×5, 7×7, and so forth. The remaining two groups to the right are more spread out but follow the same pattern. In Fig. 4(b), the groupings are not as well defined in τ space, but the patterns may still be observed.

Patterns may also be observed in the appearance and disappearance of traces. In Fig. 4(a) the source is approaching the receivers, so a ray path j appears (''cuts on'') when the range falls below the appropriate $R_{j\,\max}$ and disappears (''cuts off'') when the range falls below $R_{j\,\min}$, the range at which the grazing angle at the bottom becomes steeper than the critical angle of

$$\theta_{\mathrm{cr}} = \arccos(1500/1720) = 29°. \tag{2}$$

The appearance at $R_{j\,\max}$ thus depends mainly on the water depth and the sound-speed profile, while the disappearance at $R_{j\,\min}$ depends mainly on the bottom sound speed. As an example of ray cut-on, we observe in Fig. 4(a) that the three-traversal ray first arrives at receiver 2 near min 8 because the

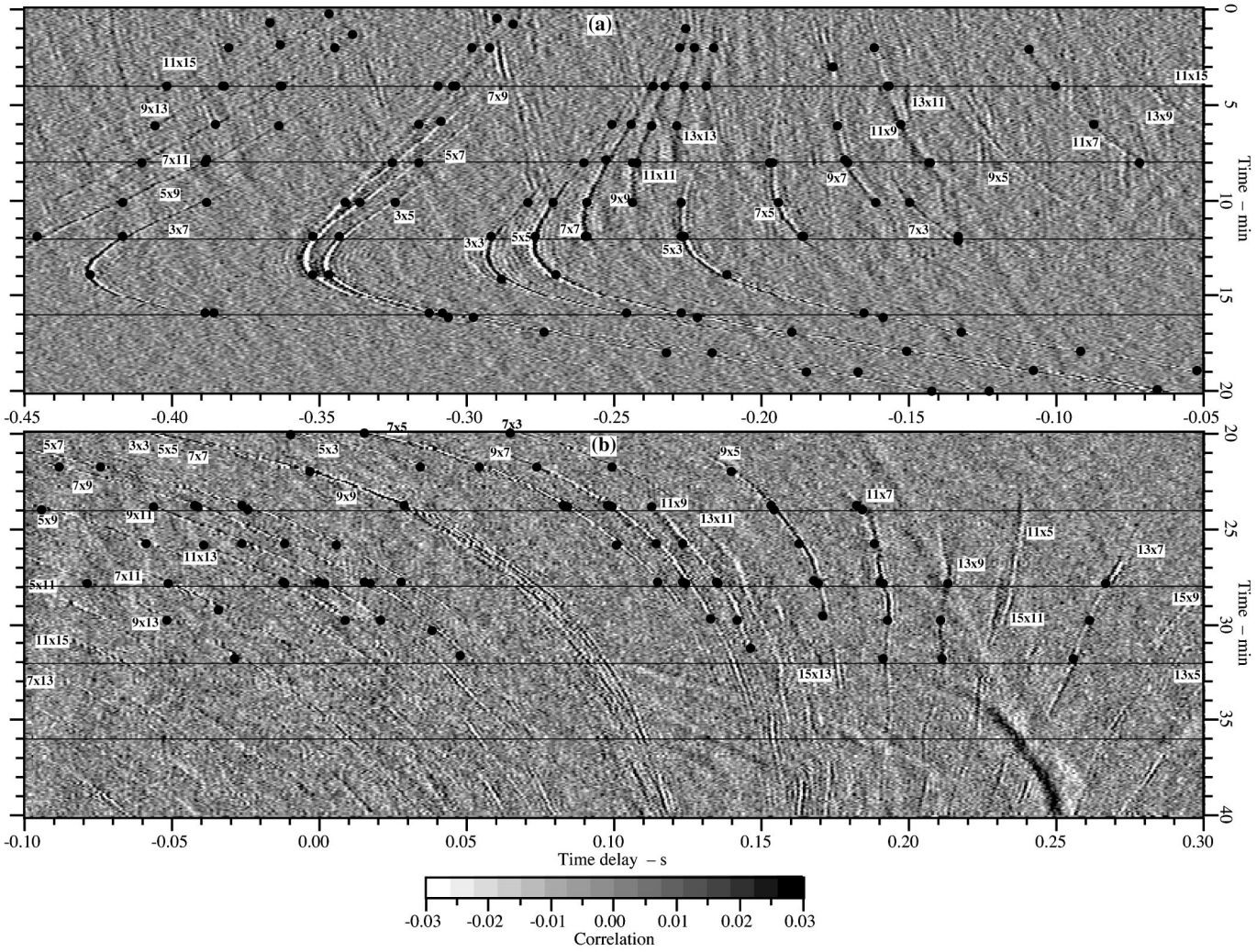


FIG. 4. Detailed view of correlogram using 10–600-Hz frequency band: (a) min 0–20; (b) min 20–40. Correlogram traces are labeled according to the ray paths at each receiver that produce them. Dots indicate sample (t,τ) points used in localization algorithm.

7×3, 5×3, and 3×3 all appear at that time. As an example of ray cut-off, we observe that the seven-traversal ray becomes too steep at receiver 1 near min 12 because the 7×7, 7×5, and 7×3 all disappear at that time.

Plots of the $\overline{t_j}(R)$ functions for ray paths with 1–19 ocean traversals for the English Channel environment are shown in Fig. 5. Dots indicate rays that interact with the bottom at the critical angle and thus define $R_{\min}$ for that ray path. Rays at shorter ranges are expected to have weaker amplitudes because of partial reflection from the bottom. Squares indicate the cut-off ranges $R_{j\ \max}$ for the first three $\overline{t_j}(R)$ functions.

The simplest propagation in terms of ray arrivals occurs near min 13, where the source is at endfire to the two receivers, and only six traces are observed in the correlogram in Fig. 4(a). The source localization results derived in the following section indicate that the ranges to receivers 1 and 2 are about 1.7 and 2.1 km, respectively. The ray plot in Fig. 1 shows the dominant eigenrays for precisely this geometry. Observe that two strong rays (the three- and five-traversals) arrive at the first receiver, whereas three (the three-, five-, and seven-traversals) arrive at the second. The six combina-

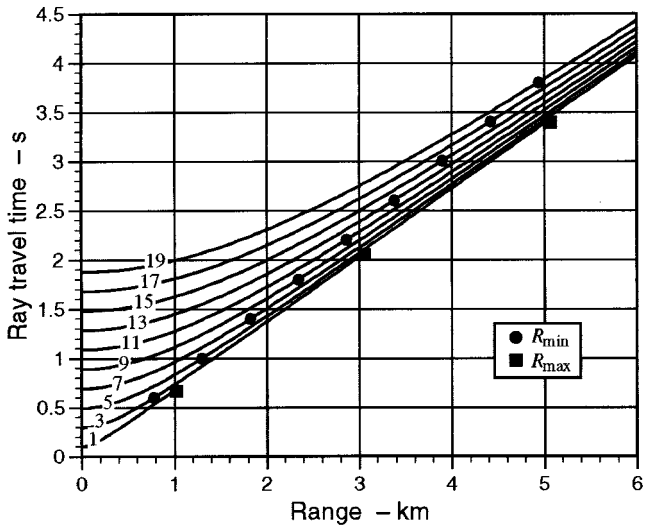


FIG. 5. $\overline{t_j}(R)$ functions for ray paths with 1–19 ocean traversals for the environment used in Sec. II. Dots indicate rays that interact with the bottom at the critical angle. Squares indicate the cut-off ranges $R_{j\ \max}$ for the first three $\overline{t_j}(R)$ functions.

tions of these eigenrays produce the six traces observed on the correlogram at min 13 in Fig. 4(a).

Careful examination of Fig. 4 also reveals that the relative positions of traces can offer a clue to the identity of groups of traces. Consider the intra-order traces 3×3, 5×5, etc., in Fig. 4(a). The higher order traces always appear toward zero time delay relative to lower order traces because they arrive at steeper vertical angles and therefore have less absolute delay. [The phenomenon is clearer in Fig. 4(a) than in Fig. 4(b) because the source is nearly at broadside in Fig. 4(b).] The higher order traces are easily identified because they cut on and cut off earlier than do the lower order traces. The two leftmost groups follow the opposite pattern and thus cannot be the intra-order group. The two rightmost groups do follow the pattern and thus cannot be ruled out on this basis. However, the traces in those groups are much too spread out to resemble the intra-order group.

C. Results of localization method

The source localization algorithm was applied in 4-min segments using the trace identifications and (t,τ) points shown in Fig. 4. A total of 234 (t,τ) points over the eight time segments was used. The lowest order ray path used was 3×3, while the highest order ray path used was 11×15.

The results of the eight independent localizations are shown in Fig. 6. The x-y locations of the start and end points of the track segments are plotted in Fig. 6(a), the range and bearing versus time of the source relative to the first receiver are given in Fig. 6(b), and the heading and speed of the source are given in Fig. 6(c).

Since the source track was not independently measured, it is not possible to determine the absolute accuracy of the localization. However, there are several indications that the accuracy is reasonably good. First, the source track segments in Fig. 6(a) are nearly contiguous. Since each 4-min segment is localized independently, there is no specific constraint that the total source track be smooth. Second, the source speed and heading [see Fig. 6(c)] are fairly constant, as would be expected for a passing ship. Variable winds or obstacles (including the experimental apparatus itself) could have caused the actual ship speed and heading to deviate from constant. Errors in water depth, errors in sound speed profile, or nonconstant bathymetry could cause localization errors. Finally, the standard deviation of the error between the measured and simulated time delays ranged from 0.86 to 1.95 ms, which is very small. Overall evidence of the good fit is shown in the following section by good agreement between simulated and measured correlograms.

An additional indication of reasonably accurate source track localization was obtained in the process of deriving the *receiver* separation. The separation was experimentally measured as 437.5 m, with an uncertainty of 150 m produced by GPS errors, deployment procedures, and ocean current. To derive the receiver separation, a separate algorithm was developed that included the second receiver's x-y position in the list of unknowns. The 12–16-min period was used to simultaneously solve for source track and receiver position. The derived receiver positions were then used to localize the

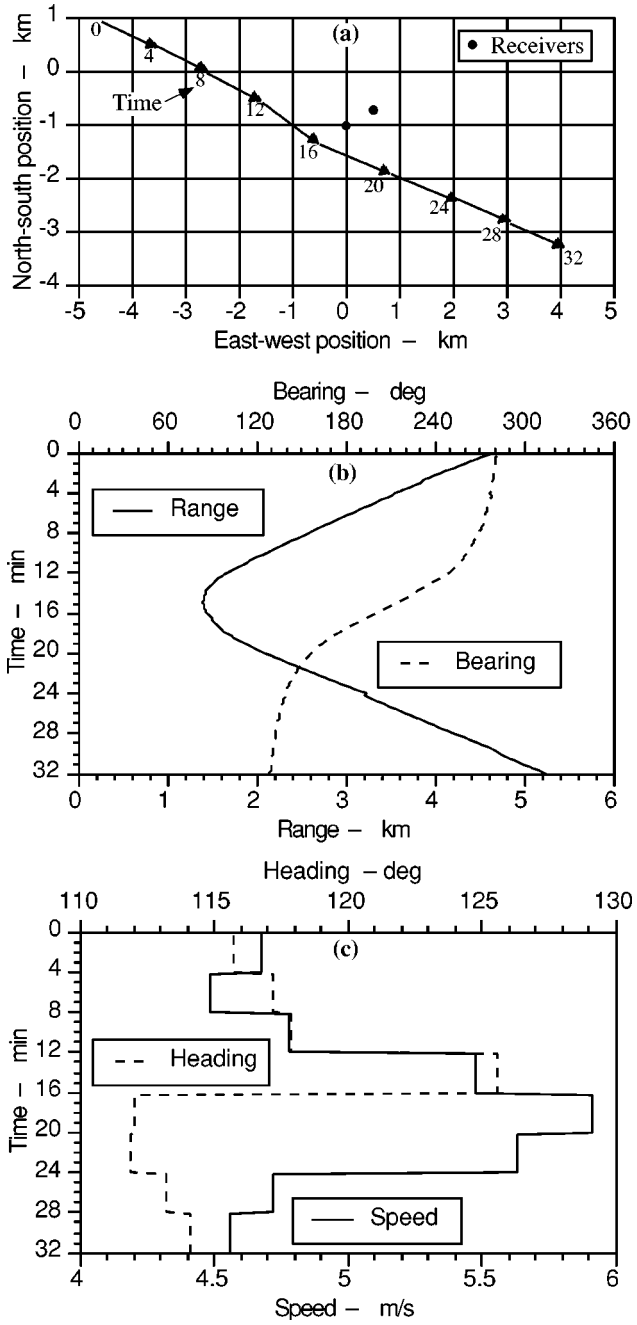


FIG. 6. (a) Track segments derived from localization algorithm. (b) Range and bearing of localized source versus time. (c) Derived source speed and heading of localized source versus time.

source over the other time periods. The derived separation was 450.2 m, which is 12.7 m, or 2.9%, different from the experimentally estimated value.

Another point regarding the localizations is that with only two receivers in a range-invariant environment, a front–back ambiguity in the source localization exists. The source localization algorithm in that case tends to converge to the solution closest to the initial guess. In our case, a third receiver, which was non-collinear to the other two, was used to break the symmetry of the two-receiver problem. The algorithms we have developed can accommodate any number of receivers, both for source localization alone and for simultaneous source and receiver localization. To keep the analysis

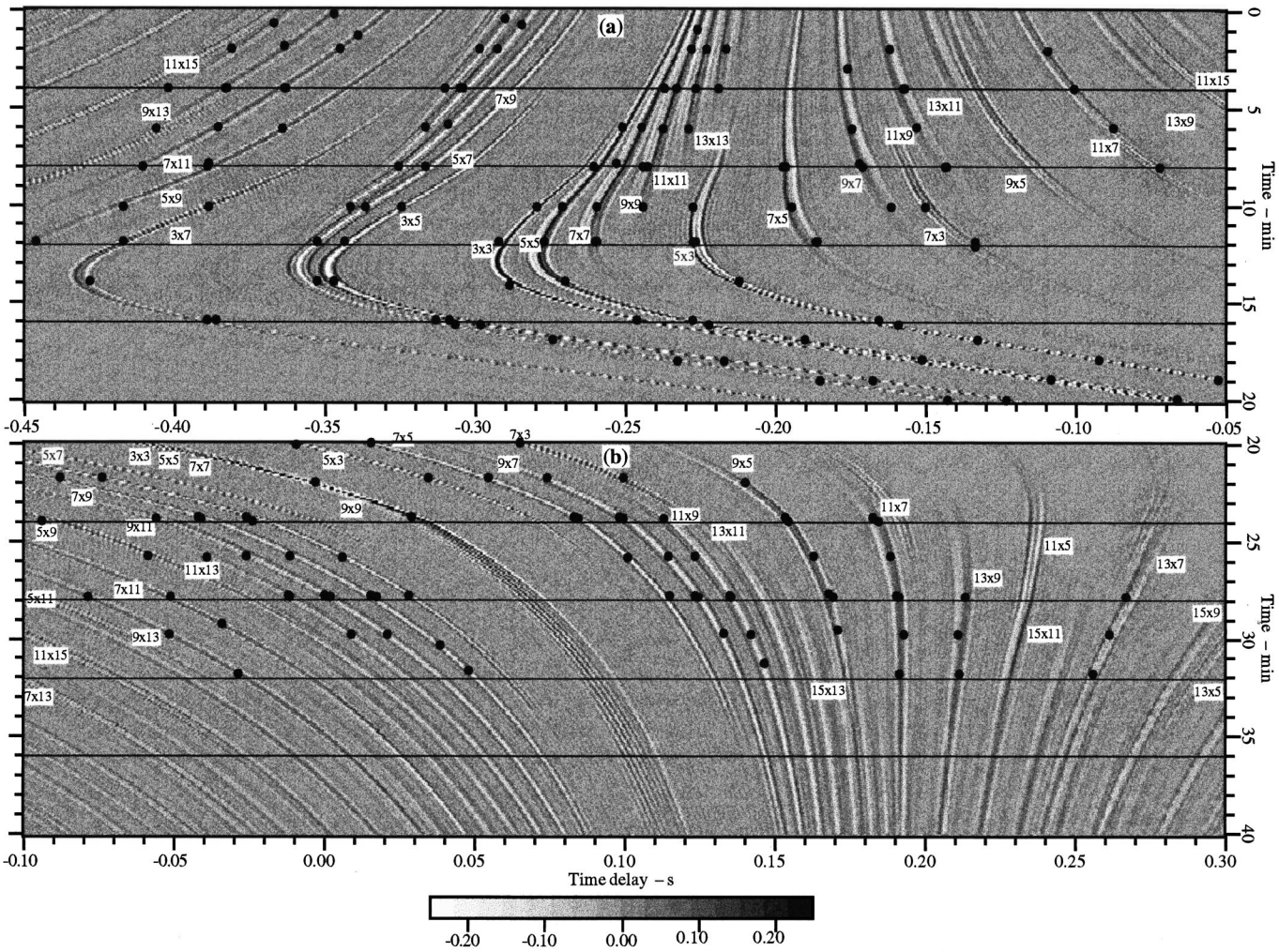


FIG. 7. Simulated correlogram using the source track derived from the localization algorithm: (a) min 0–20, (b) min 20–40. The (t,τ) points (dots) used in localization algorithm are superimposed. Compare to the correlogram of the measured data in Fig. 4.

and display of data simple, however, we have focused on the two-receiver problem in this application.

The sensitivity of the localization results to the initial guess for the source track parameters was examined by randomly varying the initial guess values. Uniform distributions of v over the interval (1,10) m/s, ϕ over (0°, 360°), and x_0 and y_0 over $(-3,3)$ km (at $t_0=12$ min, the time of CPA) were sampled. Of the 100 optimization runs made for the 16- to 20-min period, 95 resulted in the correct solution. The other five had speeds over 12 m/s and errors over twice as large as that of the correct solution. The halting of the optimization at these points, however, appears to be an artifact of the particular algorithm because, when the algorithm was continued with the same parameters, it moved away from the false solution and found the correct one. An average of 103 iterations were required for convergence over the 100 optimization runs.

The English Channel data set provides a good example of the propagation characteristics for which the localization algorithm may be applied. From Fig. 6(b), the source range is less than 2.5 km (17 water depths) over the interval 8–22 min, and the task of picking correlation traces is not difficult because the traces are quite strong and not excessively numerous (see Fig. 4). From min 22 to min 32, the range increases to just beyond 5 km (38 water depths), and trace detection and identification become more difficult. After min 32, the long range and large number of rays cause the traces to be weaker and more numerous, making the localization method difficult to apply. For this site, then, the method is applicable for ranges less than about 40 water depths.

D. Simulation of data

Figure 7 shows the simulated correlogram for the source track derived from the localization algorithm. The ORCA normal mode model[15] was run for the geoacoustic environment given previously. The model generated broadband transfer functions over the 10–600-Hz band, and the simulated spectra were processed in the same way as were the data, except that spectral whitening was not needed. The lack of noise in the simulations causes higher correlation levels (compare gray shade scales) and somewhat more traces to be observable.

Overall, excellent agreement may be observed between the measured and simulated correlograms in Figs. 4 and 7, respectively. The ray paths and the measured (t,τ) points are

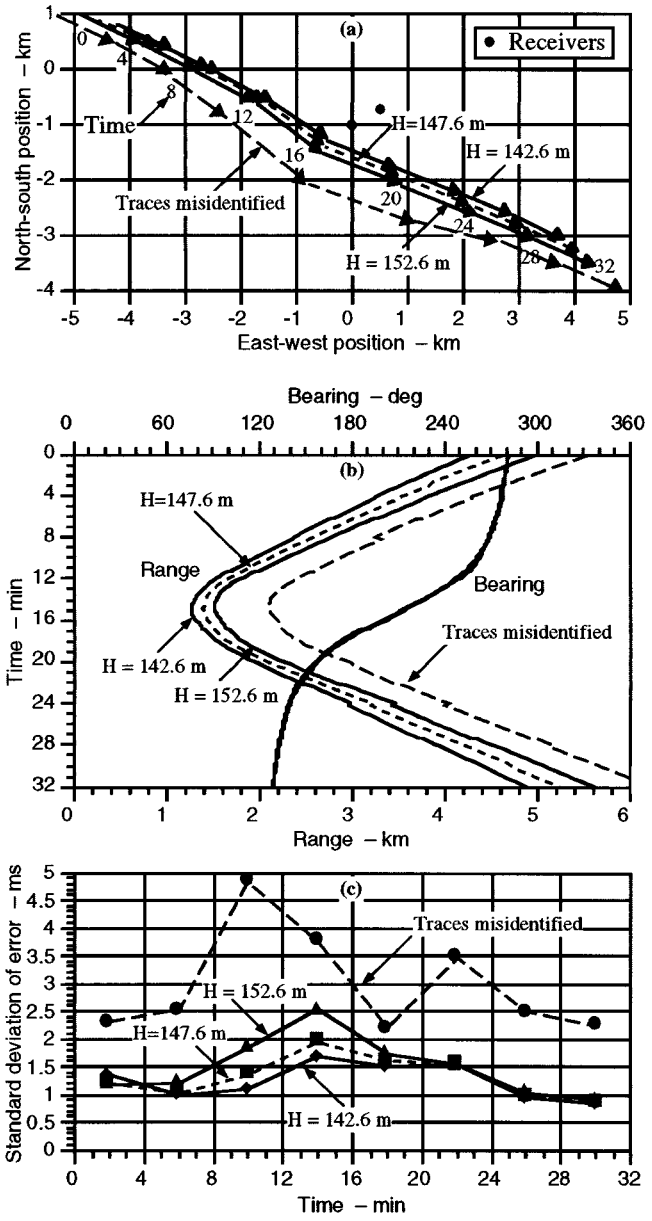


FIG. 8. Effects on source localization of (1) misidentification of ray paths (broad dashed lines) and (2) errors in water depth (solid lines): (a) derived source tracks, (b) range and bearing of source versus time, and (c) standard deviation of error between simulated and measured time delays versus time segment. The fine dashed lines correspond to the reference solution.

overlaid on the simulated correlogram to aid in the comparison. While the good fit to the (t,τ) points in the simulation is a result of the minimization of the time delay errors, it is also encouraging that the simulated correlogram does not exhibit any additional traces that are not present in the data.

III. SENSITIVITY TO MISMATCH

In this section we examine several sources of error that can arise using the localization method. Figure 8 shows plots of (a) the source tracks, (b) range and bearing versus time, and (c) the standard deviation of the error between the measured and simulated values of time delay that result from

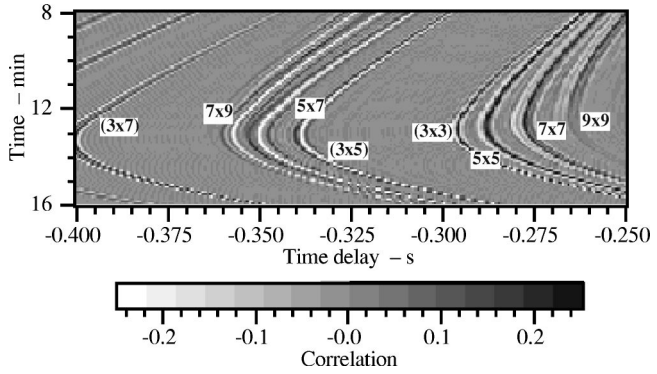


FIG. 9. Simulated correlogram using source track derived from misidentification of ray paths. Comparison with Fig. 7 reveals that the (t,τ) sample points are reasonably well matched, but additional traces involving the three-traversal ray are present.

localizations performed with various types of mismatch. The figure will be described and analyzed in the following subsections.

A. Trace identification

As noted earlier, identification of correlation traces in environments where significant multipath propagation occurs can be difficult. Traces are most easily identified near CPA, where the fewest ray paths arrive and the signal is strongest. For the correlogram in Fig. 4, the most reasonable error to make would be to add or subtract one bottom reflection (2 ocean traversals) from each ray path. For example the 3×3 trace could be mistaken for the 5×5 path or the 1×1 path. In this section we examine the effect of trace misidentification.

We first examine the effect of associating each correlogram trace with rays having one *additional* bottom reflection. The broad dashed lines in Fig. 8 show the derived track, the range and bearing versus time, and the error when the traces are misidentified. Figure 9 shows the simulated correlogram produced by the derived track for min 8–16, with the traces labeled according to their mismatched values (compare to Fig. 7). It is interesting that the optimization algorithm finds track segments that match the measured time delays reasonably well. The standard deviation of the error, however, is about twice as large for the mismatched ray paths [see Fig. 8(b)]. In addition, the overall track derived using mismatched ray paths is less straight and has more variable speed. The error and variable velocity are both worst near CPA. However, perhaps the clearest evidence that the track is not correct is provided by the simulated correlogram: Additional traces involving the three-traversal rays appear in the simulated data but not in the measured data (compare Figs. 9 and 4).

When the ray paths are misidentified by *subtracting* a bottom bounce from each path, the localization algorithm does not converge. The reason is that the only geometry for which the 1×1 trace lies near $\tau=-0.29$ s (around min 13) is one where the source is at endfire near the maximum range for the direct-path ray (0.74 km from the midpoint of the receivers). For that geometry, however, the 3×3 trace arrives at $\tau=-0.26$ s and the 5×5 trace arrives at

$\tau = -0.21$ s, time delays that are far different from the measured values of $\tau = -0.275$ s and $\tau = -0.26$ s, respectively. When the 3×3 or 5×5 traces are well matched, the source is pushed beyond the maximum range for which the 1×1 arrives.

Other trace misidentifications that were tested resulted in non-convergence. Over the 12–16-min period, the 3×3 cannot be assigned to the actual 3×5 or 5×7 traces (see Fig. 4) because the absolute time delays are too large given the receiver separation. When the 3×3 is assigned to the actual 5×3 trace and other traces are assigned in the same direction, an error of 56 ms and an unreasonable speed of 18 m/s are derived. When the relatively minor error of switching the 5×7 and 3×5 traces is made, the derived track is very similar, but the error increases to 5 ms.

In summary, ray path identification can be the most difficult part of the localization procedure, but misidentification typically causes non-convergence (characterized by extremely large errors) or impossible time delays. The only misidentification that produces a reasonable source track involves adding a bottom bounce to each ray path. In that case, however, the time delay errors are significantly larger, the total source track is more non-uniform, and additional traces are present on the simulated correlogram. This phenomenon suggests that penalizing derived source tracks that produce traces *not* present in the measured data would be beneficial.

B. Water depth and sound-speed profile

In this section we study the effect of making an error in the water depth. To do so, we simply vary the water depth by ±5 m in our ray model, produce different tables of ray travel time versus range, and use these (mismatched) tables in the localization algorithm. Although we could perform this study using (t,τ) points taken from a *simulated* correlogram in place of the experimental correlogram, we have chosen to continue to use the original data because (1) the effects should be the same and (2) we can examine the possibility that an error in water depth was indeed made. When the water depth is changed and the previous localization results are used as initial guesses, the localization algorithm converges very quickly.

The results of the localization algorithm for water depths of $H = 152.6$ m (5 m more than nominal) and $H = 142.6$ m (5 m less than nominal) are shown by the solid lines in Fig. 8. From Fig. 8(a) and (b) we observe that an increase (decrease) in the water depth causes the source range to increase (decrease) but the bearing to remain constant. (The same dependence of source range on water depth mismatch was observed in Refs. 16 and 17 using matched field and matched mode processing on a vertical aperture.) The variation of 5 m out of the 147.6-m total water depth causes the source range to vary 100–300 m, the larger values occurring at the longer ranges. The error in time delay in Fig. 8(c) is relatively small for all depths but increases slightly for the larger water depth and decreases slightly for the smaller water depth, suggesting that the water depth might be smaller than reported.

The sensitivity of the algorithm to errors in sound speed profile was tested by producing tables showing ray travel time versus range assuming a constant 1500 m/s profile instead of the profile plotted in Fig. 2. Surprisingly, the localization that resulted was indistinguishable from the one using the measured profile. In fact, the worst range difference was less than 40 m and occurred at long ranges, while the worst bearing difference was 1.5° and occurred near CPA. Even when the sound speed was set to a constant value of 1516 m/s (the speed at the surface in Fig. 2), the worst range difference increased to just 95 m, and the worst bearing difference increased to just 6°.

The robustness of the source localization with respect to water depth and sound speed profile is a positive aspect of the algorithm. It is desirable that small environmental mismatch *not* cause a localization algorithm to fail totally, but rather to degrade the solution by a small amount. For the environment studied here, errors in water depth systematically affect the range of the source but not the bearing, while reasonably large errors in sound-speed profile have little effect on the localization.

IV. SUMMARY

In this article we have developed and applied a source localization method that relies on matching measured and simulated broadband correlation traces over a period when the source is assumed to have constant velocity. The method is applicable when the source emits continuous broadband noise and when the environmental propagation characteristics allow the identification of correlation traces in terms of specific ray paths. The localization algorithm was successfully applied to measurements in a range-invariant, shallow-water area in the English Channel, where numerous ray multipaths produced a very complex correlogram. The method was shown be highly robust against errors in the water depth and sound-speed profile.

A positive aspect of the method presented here is its computational simplicity. The data processing involves a single cross correlation and a single propagation model run, the latter to obtain the $\overline{t_j}(R)$ functions for each ray path involved. Localization over each time segment then requires several hundred or less iterations over source track, each of which involves evaluating cubic spline fits of $\overline{t_j}(R)$ for each (t,τ) sample point used. Broadband matched field processing for a horizontal pair of receivers, on the other hand, would involve a large number of propagation model runs: one for each combination of source point on a fine grid in the horizontal plane and receiver point, and, for each combination, runs at a large number of frequencies would be required. Besides the model calculations, the broadband fields would need to be matched at each time interval. (For the data set examined in this article, 1208 frequency bins and 417 time intervals were used.) Finally, the output of the matched field processor (for an assumed source depth) would be a three-dimensional array of filter output versus x, y, and time. Further processing would be required to extract the source track. For arrays with only vertical aperture, the computational complexity of MFP is considerably reduced, but still quite large for the broadband problem.

The computational simplicity of the present method lies partly in the fact that source detection, which requires no

environmental information, is done before, rather than simultaneously with, localization. Source detection is achieved via broadband cross correlation and provides the ''matching space'' quantities for the present approach: ray travel time differences. The number of ray travel time differences to be matched is on the order of 2–10, whereas matching in complex pressure field space requires matching at 100–1000 frequency bins. Further reduction in computation time is achieved when the source is assumed to travel in a straight line at constant velocity over a given time segment.

The most difficult part of the localization procedure is the identification of the rays that produce the correlogram traces. It was shown that, for the English Channel data analyzed, adding a bottom reflection to each ray path was the only misidentification that produced a reasonably good fit to the observed traces. Other misidentifications resulted in very large errors between the simulated and observed time delays. Still, the major drawback to the approach is the time and analysis typically required to identify the ray paths involved. Automation of the process will require additional work, possibly along the lines of testing a set of hypothesized trace identifications to find the one that results in the least error.

Another aspect of the algorithm that may cause difficulty in certain circumstances is the assumption of a known source depth. When the source depth is small compared with the total water depth, the assumptions and procedures used in this work are applicable (see Sec. I B). This would apply to all surface ships, as well as to submerged sources in not-too-shallow water. When the source is deeper, however, the correlation trace structure changes. The downward- and upward-traveling rays for each order of ray multipath separate and produce twice as many ray paths at each receiver, causing four times as many correlation traces. Ray path identification might be more difficult, although the sign of the correlation traces could be helpful in this regard. Similarly, when the receivers are off the bottom, the number of correlation traces increases.

The last area of potential difficulty lies in identifying the type of bottom-interacting ray path that dominates. The high sound speed (1720 m/s) of the fine sand present in the English Channel area made the choice of specular reflection easy to make. The localization method has also been successfully applied to an area in the Mediterranean Sea, where the bottom-penetrating and -turning rays dominate the specular reflection. Especially at short range, use of the correct bottom-interacting ray path was found to be important.

Extension of the localization method to range-variable environments would involve tabulation of the $\overline{t_j}(R)$ functions for a set of radials using a range-dependent ray model. Computation of the ray travel times for a particular source location would then require interpolation in azimuth. The complexity of the code would increase, but the overall computation time would not increase significantly.

The method may be applied to any type of receiver configuration. Ray multipath at single receivers can be used to localize a source (with no bearing information, of course). More than two receivers may be used by choosing (t,τ) points from multiple correlograms and accounting for the receiver pairs involved with each correlation trace. Vertical aperture may also be used by computing and using the $\overline{t_j}(R)$ functions corresponding to each receiver water depth.

ACKNOWLEDGMENTS

This work was supported by ONR 321. The authors also wish to acknowledge Karl Focke for his contributions in collecting the data and in providing details of the experiment.

[1] J. C. Hassab, ''Passive tracking of a moving source by a single observer in shallow water,'' J. Sound Vib. **44**, 127–145 (1976).
[2] D. H. McCabe and R. L. Moose, ''Passive source tracking using sonar time delay data,'' IEEE Trans. Acoust. Speech Signal Process. **29**, 614–617 (1981).
[3] R. L. Moose, and T. E. Dailey, ''Adaptive underwater target tracking using passive multipath time-delay measurements,'' IEEE Trans. Acoust. Speech Signal Process. **33**, 777–787 (1985).
[4] H. C. Schau and A. Z. Robinson, ''Passive source localization employing intersecting spherical surfaces from time-of-arrival differences,'' IEEE Trans. Acoust. Speech Signal Process. **35**, 1223–1225 (1987).
[5] J. S. Abel and K. Lashkari, ''Track parameter estimation from multipath delay information,'' IEEE J. Ocean Eng. **12**, 207–221 (1987).
[6] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994).
[7] B. H. Maranda and J. A. Fawcett, ''Detection and localization of weak targets by space-time integration,'' IEEE J. Ocean Eng. **16**, 189–194 (1991).
[8] B. H. Maranda and J. A. Fawcett, ''Localization of a maneuvering target using simulated annealing,'' J. Acoust. Soc. Am. **94**, 1376–1384 (1993).
[9] J. A. Fawcett and B. H. Maranda, ''A hybrid target motion analysis/matched-field processing localization method,'' J. Acoust. Soc. Am. **94**, 1363–1371 (1993).
[10] E. K. Westwood, ''Element Localization of the LFA-5 VEDABS Arrays,'' Applied Research Laboratories Technical Report No. 92-3 (ARL-TR-92-3), Applied Research Laboratories, The University of Texas at Austin (1992).
[11] P. E. McCarty and E. K. Westwood, ''Array Element Localization of the CST5 and LFA7 VEDABS Arrays,'' Applied Research Laboratories Technical Report No. 92-20 (ARL-TR-92-20), Applied Research Laboratories, The University of Texas at Austin (1992).
[12] E. K. Westwood and P. J. Vidmar, ''Eigenray finding and time series simulation in a layered-bottom ocean,'' J. Acoust. Soc. Am. **81**, 912–924 (1987).
[13] E. K. Westwood and C. T. Tindle, ''Shallow water time series simulation using ray theory,'' J. Acoust. Soc. Am. **81**, 1752–1761 (1987).
[14] MATLAB *High Performance Numeric Computation and Visualization Software User's Guide*, The Math Works, Inc., Natick, MA (1992).
[15] E. K. Westwood, C. T. Tindle, and N. R. Chapman, ''A normal mode model for acousto-elastic ocean environments,'' J. Acoust. Soc. Am. **100**, 3631–3645 (1996).
[16] D. R. Del Balzo, C. Feuillade, and M. M. Rowe, ''Effects of water-depth mismatch on matched field localization in shallow water,'' J. Acoust. Soc. Am. **83**, 2180–2185 (1988).
[17] E. C. Shang and Y. Y. Wang, ''Environmental mismatching effects on source localization processing in mode space,'' J. Acoust. Soc. Am. **89**, 2285–2290 (1991).

Noise directionality for surface sources in range-dependent environments

C. H. Harrison
BAeSEMA, Apex Tower, New Malden, Surrey KT3 4LH, England

(Received 6 November 1996; revised 23 April 1997; accepted 30 June 1997)

An earlier paper [J. Acoust. Soc. Am. **99**, 2055–2066 (1996)] tackled the calculation of ambient noise and noise coherence in a range-independent reflecting or refracting environment with horizontal distributions of noise sources. The approach was analytical producing some closed-form solutions and never more than a single well-behaved integral to solve. In this paper the azimuth and elevation dependence of the noise over a sloping bottom is investigated. The analytical and graphical results explicitly show the effects of upslope and downslope bottom reflections on the ''noise notch,'' and in passing they explain the flatness of the angle distribution sought by Dashen and Munk [J. Acoust. Soc. Am. **76**, 540–554 (1984)]. It is also clear that, contrary to one's intuition, three-dimensional horizontal bending effects on the noise directionality may be significant because of the large contribution from distant sources. © *1997 Acoustical Society of America.*
[S0001-4966(97)01111-9]

PACS numbers: 43.30.Nb, 43.30.Es, 43.30.Cq [SAC-B]

INTRODUCTION

The relationship between ambient noise, noise mechanisms, sound propagation, and the underwater environment has been intriguing for a long time. This is partly because an understanding of noise coherence and its associated directionality is essential for the design of sonar beam patterns and signal processing. It is also because there are still many unanswered physics questions to do with the detailed mechanisms of wind, wave and rain noise generation, noise statistics, and spectra. On top of this there are the problems of summing the propagation paths from many distributed noise sources to one receiver, and this is the subject on which we shall concentrate.

Cron and Sherman's[1] early analytical noise model assumed distributed sound sources but excluded bottom reflections. Since then most developments have been numerical and can therefore include much more environmental detail.[2–7] A recent review of models is given by Hamson.[8] Models of shipping noise, such as RANDI,[2,3] usually sum the contributions from a finite number of discrete sources. In contrast, Perkins *et al.*[9] have developed a model with three-dimensional capability that tackles distributed sources. This solution is based on precomputed, stored local modal data[10] and an extended version of Kuperman and Ingenito's noise model[11] that uses adiabatic mode theory.

In a range- and azimuth-dependent environment there is at least one important effect that is not seen in a stratified medium. This is the well-known diversion of steeply reflected paths into the sound channel. In this way the angle range in deep water that would otherwise constitute the ''noise notch''[12,13] may be filled in by contributions from distant shipping or wind noise. Often one may see a flat-topped angle distribution under these conditions, as reported by several authors.[12–16] Dashen and Munk[17] attempted to construct a model of this behavior but came to the conclusion that there was ''some mystery as to how this [mechanism] will lead to the observed high degree of mode equipartition,'' i.e., a flat-topped ''pedestal'' in angle. Sotirin and Hodgkiss[18] have more recently reviewed the experimental and modeling work, and they carried out a detailed investigation of the statistics, spectra, and angular response of a high resolution vertical array, with this problem, among other things, in mind.

The main aim of this paper is to investigate the noise directionality that results from a uniform distribution of surface sources in a range-dependent environment, such as one of those shown in profile in Fig 1. We acknowledge that numerical models such as some of those cited above could easily handle this problem, but our purpose here is to understand the functional relationships, and above all to explain the flat response in a simple way. For this reason our approach is analytical.

I. FORMULATION

An earlier paper[19] derived a formula for noise intensity and coherence, and although the subsequent examples were for range-independent media, it was noted that the original formulation was still valid in an arbitrarily range-dependent and azimuth-dependent medium (as may be verified by flux arguments). We adopt a ray approach again, confident that it has been reconciled with normal mode approaches[19] and that it can handle as much multipath interference as is necessary for dealing with a uniform distribution of surface noise sources.

An important point in this formulation for distributed sources is that the ray spreading that weakens the intensity contributions from sources on the surface is exactly compensated by the corresponding increase in number of sources in the beam. Thus the usual geometric spreading effects cancel out, and we simply have to account for boundary and volume losses.

To recapitulate, the noise sources, are assumed to be at the surface, and the (small) hydrophone separation and elevation angle are d and γ; $d_v = d \sin \gamma$, $d_h = d \cos \gamma$. Both

hydrophones are within the dot in Fig. 1. The additional azimuth offset β of the hydrophone pair now allows us to fix the bathymetry in x, y, z and still calculate the correlation function for arbitrary hydrophone orientations. We have to integrate ray contributions over all azimuths ϕ and all elevation angles θ_r (at the receiver); locally ray elevation angles are related by Snell's law. We also have to add paths that have had a number of surface and bottom reflections. In the range-independent case this was particularly straightforward, but in general this is not so.

From Eqs. (7) and (8) in Harrison[19] we can write the (unnormalised) correlation function (for unit source strength per unit area) as

$$\rho(d,\beta,\gamma)=\int_0^{2\pi}\int_{-\pi/2}^{\pi/2} e^{ikd_h\cos(\phi-\beta)\cos\theta_r} e^{ikd_v\sin\theta_r} \times D(\phi,\theta_r)\cos\theta_r\, d\theta_r\, d\phi, \quad (1)$$

where k is wave number at the receiver, and the noise directionality D is the product of two terms

$$D(\phi,\theta_r)=Q(\theta_r)S(\phi,\theta_r). \quad (2)$$

The function Q is the residual attenuation due to bottom reflection (power reflection coefficient R_b) and volume absorption (a) between the last ray upper turning point and the receiver. For the upward ray the partial path length is s_p, and for the downward path it is (s_c-s_p), where s_c is the complete cycle path length. So for a path steep enough to hit the surface, Q is given by

$$Q(\theta_r)=e^{-as_p}, \quad \theta_r\geqslant 0,$$
$$=R_b(\theta_b)e^{-a(s_c-s_p)}, \quad \theta_r<0. \quad (3)$$

Central to this paper is the function

$$S(\phi,\theta_r)=\sum_{n=0}^{N}\left(\prod_{j=1}^{n}(\Re_j)\right)\sin^{2m-1}(\theta_s)_n, \quad (4)$$

with

$$\Re_j=R_s((\theta_s)_j)R_b((\theta_b)_j)\exp(-a(s_c)_j). \quad (5)$$

Here, R_s and R_b are treated as symmetrical functions of angle.

Here, $S(\phi,\theta_r)$ represents the sum of contributions arriving at (ϕ,θ_r) from all ray paths connecting a surface noise source to the receiver. Generally the ray from a noise source may have n surface reflections on the way to the receiver and there are therefore n full ray cycles of attenuation. We indicate the cycle number (working outward from the receiver) by the added subscript j, and each cycle has an attenuation $\Re_j$, composed of local surface loss $R_s(\theta_s)$, bottom loss $R_b(\theta_b)$, and volume absorption $\exp(-as_c)$. The last term in Eq. (4) contains the noise source directionality which for a dipole has $m=1$. The product in Eq. (4) therefore represents the cumulative attenuation along the paths out to the nth surface interaction, and the sum represents the addition of contributions from rays with different numbers of surface interactions n. Note that the meaning of $\Pi_{j=1}^{0}$, which we always obtain for the direct path is simply unity (because for this path there is no complete ray cycle). The residual attenuations are taken care of by Eq. (3), and the usual spreading losses are taken care of in the derivation of Eq. (1) (see Harrison[19]). In the case of a range-independent environment $\Re_j$ is a constant ($\Re_j=\Re$) for each θ_r and so S reduces to the geometric series whose solution is

$$S=\sin\theta_s(1-\Re^{N+1})/(1-\Re). \quad (6)$$

In the range-dependent case $\Re_j$ may change from cycle to cycle as ray angles at the surface and bottom and cycle distances change. Because this depends above all on bottom topography, it is implicitly dependent on azimuth as hinted in Eq. (4).

Thus there are two main problems to be tackled in calculating the noise coherence. One is the sum over multipaths, i.e., the solution of Eq. (4). The other is solution of the resulting azimuth integral in Eq. (1). Here we find that the behavior of the noise directionality D [mainly through S in Eq. (4)] is interesting in its own right and the rest of the paper is devoted to its behavior alone. We leave the problem of integration in azimuth to another occasion.

II. SUM OF MULTIPATH REFLECTIONS

In this section we evaluate Eq. (4). We can visualize the changing terms in the sum and product with the aid of Fig. 1, where we have a central receiver and a generally deepening environment toward the left. Figure 1(a), (b), (c), (d) assumes respectively, isovelocity, uniform negative sound speed gradient, bilinear sound channel, and uniform positive gradient. Upper and lower ray turning points can be reflecting or refracting, and generally for a particular value of θ_r at the receiver we can have transitions between regions of surface/bottom reflection, surface duct, bottom duct, and sound channel. However, it is easy to see that not all transitions are possible if we merely change bathymetry; we will return to this later. In contrast, the assumption of constant depth means that there are no transitions, and one can completely ignore paths that do not reach the surface since they can never contribute any noise from surface sources.

In a mildly range-dependent environment it is well known that the behavior of ray elevation angle can be described through an invariant[20]

$$T=\int \frac{\sin\theta}{c}\, dz. \quad (7)$$

This is most useful when calculating ray trajectories,[21] but here we are interested merely in ray angles for the jth reflection without regard to where those angles occur in range or depth [see Eq. (3)]. For instance, in isovelocity water with a constant bottom slope ϵ the downslope ray angle θ_j at the jth reflection is simply incremented from its initial value θ_0 to $\theta_j=\theta_0+j2\epsilon$. This is enough to define Eq. (4) uniquely although its direct solution is not obvious.

For azimuths other than direct up or down slope there is an additional horizontal bending effect[21] which could be tackled using ray invariants. However, initially we assume that these effects are only slight with lossy boundaries, so we choose to ignore them. Later we will show that they are not

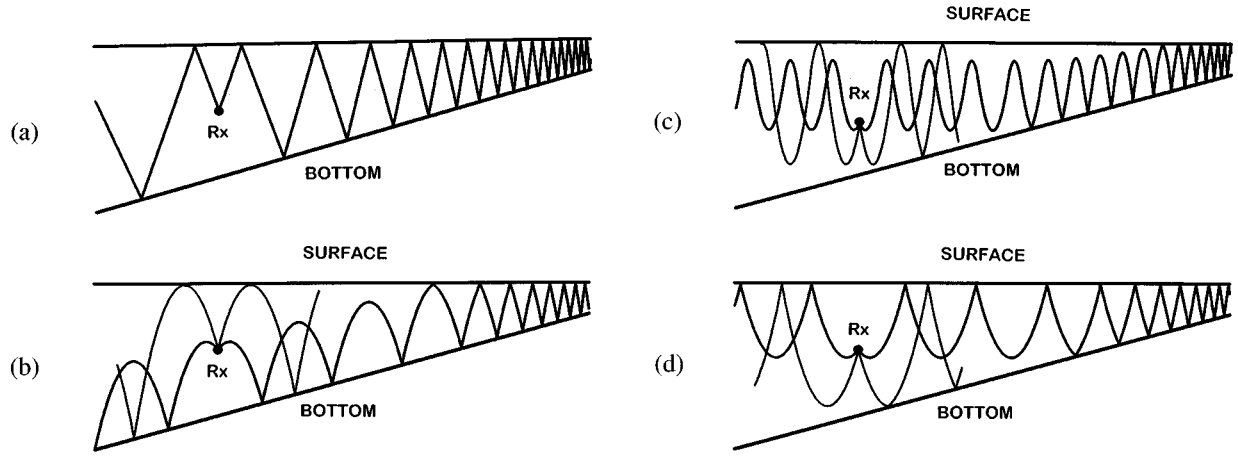


FIG. 1. Bathymetric profile and characteristic rays arriving at a receiver (Rx) for (a) isovelocity (SB), (b) negative velocity gradient (BD), (c) bilinear (SC), (d) positive velocity gradient (SD).

necessarily negligible. Instead, we take an ''$N\times 2D$'' approach taking the gradient in the appropriate bottom profile for the azimuth under consideration.

A. A theorem for multiple contributions

As Eq. (4) stands it is to all intents and purposes intractable, but we can assume that the angle changes between bounces are small enough to be represented as a continuum. This is equivalent to the adiabatic approximation as applied to normal mode theory and leads to a number of representations. First one can write $\Re_j$ (the power attenuation) as

$$\Re_j = \exp(-L_j), \tag{8}$$

where L_j is the complete boundary and volume loss for one cycle (measured in half-nepers). Then we write $\Sigma_j \rightarrow \int dj$, and Eq. (4) (for a dipole, $m=1$) becomes

$$S(\phi,\theta_r) = \int_0^N \sin\,\theta_s e^{-\int_0^n L\,dj}\,dn. \tag{9}$$

Remember that n is the number of complete ray cycles between the noise source and receiver. Its maximum possible value is N. The variable j is the identifier used to sum the losses out to cycle n.

Now, there is one interesting case for which Eq. (9) can be solved without invoking ray invariants. If we assume the dominant loss is proportional to $\sin\theta_s$ so that $L=\alpha_s \sin\theta_s$, we can write Eq. (9) as

$$S = \frac{1}{\alpha_s}\int_0^N \alpha_s \sin\,\theta_s e^{-\int_0^n \alpha_s \sin\,\theta_s\,dj}\,dn. \tag{10}$$

In isovelocity water or when rays are steep the distinction between surface and bottom angles disappears and the surface subscript s can be dropped. It is easy to verify that the integrand is a perfect differential, so that the solution of the integral is simply

$$S = -\frac{1}{\alpha_s}\left[e^{-\int_0^n \alpha_s \sin\,\theta_s\,dj}\right]_0^N, \tag{11}$$

which leads to the theorem

$$S = \frac{1}{\alpha_s}\left(1 - e^{-\int_0^N \alpha_s \sin\,\theta_s\,dj}\right). \tag{12}$$

It is also possible to obtain solutions of Eq. (10) for other assumed L dependence by integrating by parts, but this will not be pursued here. Note that the derivation of Eq. (12) has not made any assumption about how or why θ_s varies from bounce to bounce.

Comparing Eq. (11) with the range-independent case [Eq. (6)] we see that it is surprisingly similar, since $1-\Re \simeq \alpha_s \sin\theta_s$ and the exponential in Eq. (12) with constant integrand is simply $\exp(-(N+1)\alpha_s \sin\theta_s) = \Re^{N+1}$. The main difference is precisely this replacement of the exponent by an integral over all reflections. In particular, if N is very large there is no difference at all between range dependent and range independent! What this means is that S does not depend on the *way* θ_s varies since steepness of ray and high loss are compensated by contributions from a large number of cycles.

A moment's thought, however, demonstrates that N cannot always be large. For instance, going downslope (from the receiver) ray angles reduce and we reach a finite N when rays cease to interact with the bottom. Even going upslope N is governed by a critical angle and is therefore finite [although, admittedly the exponential term in Eq. (12) may tend to zero].

III. SURFACE AND BOTTOM REFLECTION IN ISOVELOCITY WATER

We can evaluate Eq. (12) by using some of the other representations alluded to earlier. One is to write the ''element'' of cycle number dj (or dn) in terms of the cycle distance and an element of range

$$dj = dr/r_c. \tag{13}$$

This was done by Weston[22] to evaluate boundary losses in range-dependent propagation. In an isovelocity medium we have $r_c = 2H\,\cot\theta$ and the integral in the exponent becomes

$$I = \int_0^n L\,dj = \alpha\int_0^r \frac{\sin\,\theta\,\tan\,\theta}{2H}\,dr. \tag{14}$$

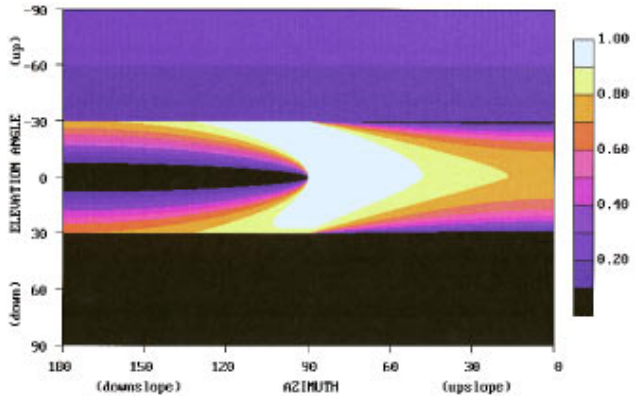


FIG. 2. Normalized noise directionality (αD) versus azimuth and elevation angle: isovelocity, $N\times 2D$, loss $\alpha=1$ dB/rad, slope=0.01, $\theta_c=30°$.

Hereafter we make no distinction between surface and bottom loss and α represents their sum. Then using the ray invariant $T=(H\sin\theta)/c$ approximated to small angles we obtain

$$I=\frac{\alpha(cT)^2}{2}\int_0^r\frac{dr}{H^3(r)}, \tag{15}$$

which can be evaluated given the bathymetric profile $H(r)$ and the ray angle at the receiver θ_r.

A. Uniform bottom slope in isovelocity water

A neater and more instructive approach for constant (small) bottom slope ϵ_0, that does not require the ray invariant can be derived from Eq. (9) by putting

$$dj=d\theta/2\epsilon, \tag{16}$$

where we have assumed a $N\times 2D$ representation with

$$\epsilon(\phi)=\epsilon_0\cos\phi \tag{17}$$

and azimuth measured relative to the upslope direction. Equation (9) becomes

$$S=\int_{\theta_1}^{\theta_2}\sin\theta e^{-\int_{\theta_1}^{\theta}\alpha\sin\theta\,d\theta/2\epsilon}\frac{d\theta}{2\epsilon}, \tag{18}$$

with θ_1 and θ_2 being initial and final ray angle, whether traveling up or down the slope. Here we can regard α as a joint surface and bottom loss per radian. This is easily evaluated using Eq. (12), or not, as

$$S=\frac{1}{\alpha}(1-e^{-(\alpha/2)|(\cos\theta_1-\cos\theta_2)/\epsilon|}), \tag{19}$$

and the modulus sign shows that the sign change in ϵ in going from up to downslope is compensated by the sign of the difference between θ_1 and θ_2. In passing, it is easy to convert Eq. (15) into Eq. (19) by putting $dr=-dH/\epsilon$.

Looking downslope from the receiver [see Fig. 1(a)] we have $\theta_1\equiv\theta_r$ and $\theta_2\equiv 0$ we obtain

$$S_{\text{down}}(\phi,\theta_r)=\frac{1}{\alpha}(1-e^{-(\alpha/2|\epsilon|)(1-\cos\theta_r)}), \tag{20}$$

whereas going upslope from the receiver [Fig. 1(a)] we have $\theta_1\equiv\theta_r$ and $\theta_2\equiv\theta_c$, where θ_c is a critical angle, we have

$$S_{\text{up}}(\phi,\theta_r)=\frac{1}{\alpha}(1-e^{-(\alpha/2|\epsilon|)(\cos\theta_r-\cos\theta_c)}). \tag{21}$$

In the case of $\theta_c=\pi/2$ we have

$$S_{\text{up}}(\phi,\theta_r)=\frac{1}{\alpha}(1-e^{-(\alpha/2|\epsilon|)\cos\theta_r}). \tag{22}$$

It is now clear that where neither S_{up} nor S_{down} can be greater than the range-independent equivalent S_{flat}, defined as

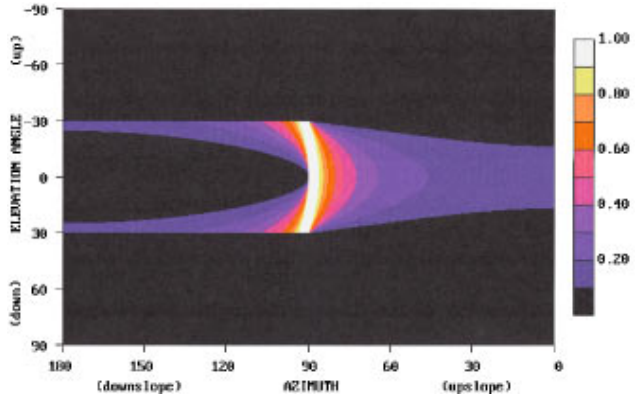


FIG. 3. Normalized noise directionality (αD) versus azimuth and elevation angle: isovelocity, $N\times 2D$, loss $\alpha=0.1$ dB/rad, slope=0.01, $\theta_c=30°$.

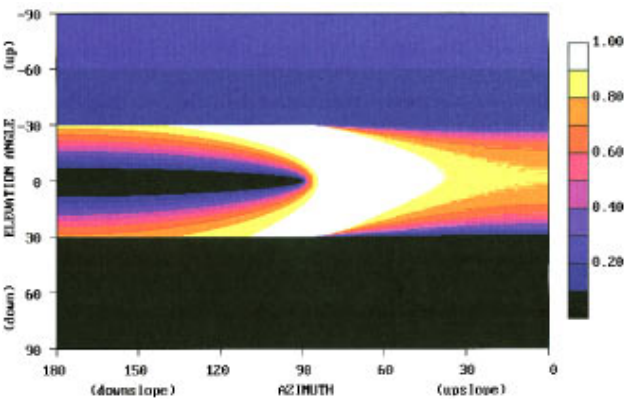


FIG. 4. Normalized noise directionality (αD) versus azimuth and elevation angle: isovelocity, 3D, loss $\alpha=1$ dB/rad, slope=0.01, $\theta_c=30°$.

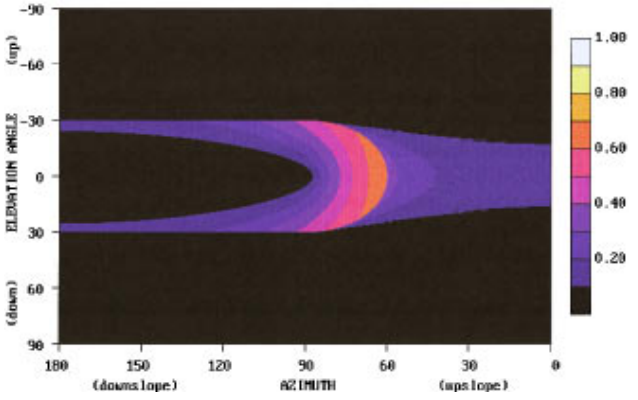


FIG. 5. Normalized noise directionality (αD) versus azimuth and elevation angle: isovelocity, 3D, loss $\alpha=0.1$ dB/rad, slope=0.01, $\theta_c=30°$.

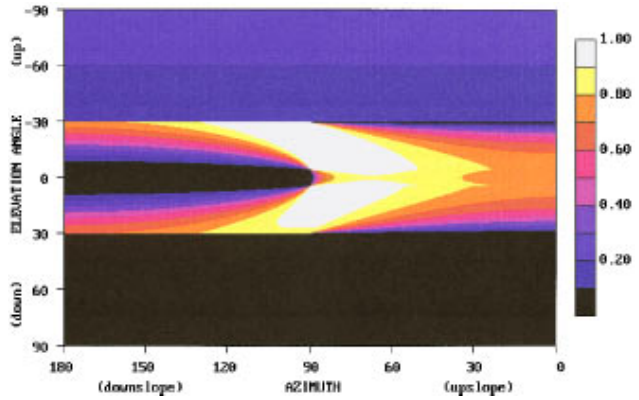


FIG. 6. Normalized noise directionality (αD) versus azimuth and elevation angle: Bottom Duct (BD), loss $\alpha=1$ dB/rad, slope=0.01, $\theta_c=30°$.

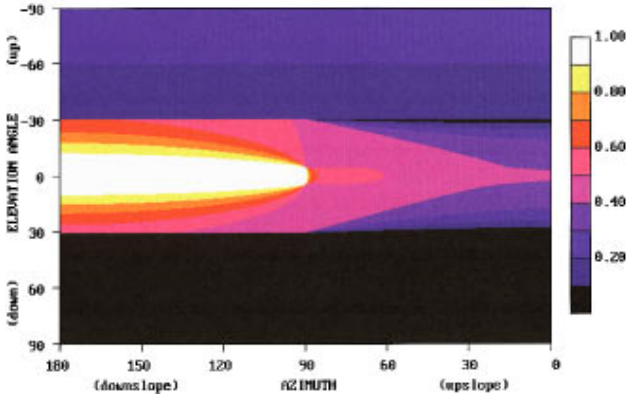


FIG. 8. Normalized noise directionality (αD) versus azimuth and elevation angle: Surface Duct (SD), loss $\alpha=1$ dB/rad, slope=0.01, $\theta_c=30°$.

$$S_{\text{flat}}(\phi,\theta_r)=\frac{1}{\alpha}. \tag{23}$$

They can only equal it for low slopes or large losses. Remembering that in a $N\times 2D$ scheme ϵ incorporates azimuth effects through Eq. (17), this means that there is always *less* noise over a slope than over the equivalent flat bottom, although the flat-bottom noise can probably be equaled in the cross-slope direction.

Intensity contours for normalised directionality ($\alpha D=\alpha SQ$) are shown in Cartesian (ϕ,θ_r) projection in Fig 2. Parameters are: Bottom slope $\epsilon_0=0.01$, bottom loss $\alpha=0.23$ (i.e., 1 dB per radian), critical angle $\theta_c=30°$. The intensity for angles steeper than θ_c at the receiver is obviously zero as indicated by the black area. The highest intensities (white) are seen slightly up-slope of across-slope. Upslope the weakest returns are for steep ray angles [Eq. (18)]; downslope the weakest returns are in the horizontal [Eq. (17)]. The up/down asymmetry is entirely due to bottom loss $R_b=\exp(-\alpha\sin|\theta_r|)$ in Q [Eq. (3)] since absorption has been set to zero. At upward elevation angles greater than critical the formula reverts to Cron and Sherman's leaving $D=\sin\theta_s$.

The plot in Fig. 3 shows the effect of reducing α/ϵ. The noise becomes *more* acutely directional for *lower* slopes or *weaker* losses because of the dependence of all formulae on α/ϵ; actually this case has changed only α to $\alpha=0.023$ (i.e., 0.1 dB per radian). Of course, the *value* of D, as opposed to αD, is higher for weaker losses.

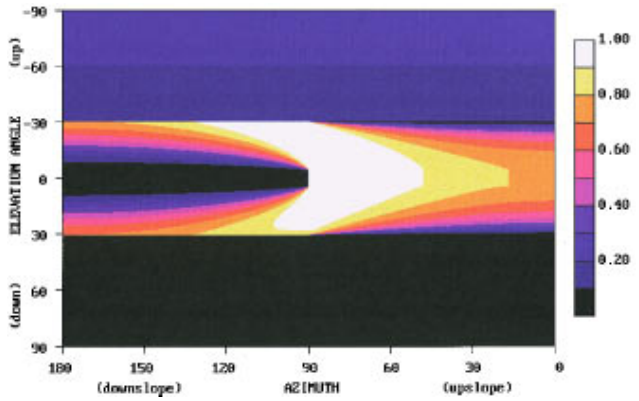


FIG. 7. Normalized noise directionality (αD) versus azimuth and elevation angle: Sound Channel (SC), loss $\alpha=1$ dB/rad, slope=0.01, $\theta_c=30°$.

B. True 3D horizontal curvature

The horizontal trajectory of a ray setting off across-slope from the receiver tends to curve towards deep water because of the minor heading change at each bottom reflection.[21] Thus the $N\times 2D$ assumption that trajectories are straight is convenient but not quite correct. The multiple reflections in the wedge can be imagined as a single straight ray passing through the surface and bottom image planes fanned out like the pages in an open book. We need the relation between grazing angle $(\theta_s)_j$ and image number (or surface reflection number) j. The angle between $(j+1)$ image planes is $2j\epsilon_0$, where ϵ_0 is the bottom slope, so the angle at the wedge apex between the receiver and the jth image plane is $(2j\epsilon_0+\delta)$, where δ is a simple constant related to receiver depth and its distance from the apex. It is then possible to construct a spherical triangle in which $(\theta_s)_j$ is the *side* opposite the initial azimuth *angle* ϕ, and the other two *sides* are the tilt of the local normal $(2j\epsilon_0+\delta)$ and the complement of initial elevation angle $(\pi/2-\theta_r)$. Using the spherical cosine rule[23] we then have the exact formula

$$\sin(\theta_s)_j=\cos\phi\cos\theta_r\sin(2j\epsilon_0+\delta)+\sin\theta_r\cos(2j\epsilon_0+\delta). \tag{24}$$

We can still do the integral in Eq. (12), but unfortuantely the upper limit N is defined by setting $\theta_{sN}=\theta_c$ in Eq. (24) which means that the closed form solution is rather messy. Instead we have the option of evaluating the integral in Eq. (12) numerically using Eq. (24) for $\theta_{sj}\leqslant\theta_c$ or reverting to Eq. (4) which can be implemented in a single loop.

Figure 4 shows intensity calculated using the former approach [in fact the latter gives almost identical results, vindicating the theorem in Eq. (12)]. Comparing this with Fig. 2 we see some slight but interesting differences mainly near the cross slope direction. There is now a larger contribution from rays that start slightly upslope but potentially extend to infinity into deep water. The rippled effect on the right of Fig. 4 is due to sudden ray truncation effects resulting from the approach just described.

Figure 5 shows an interesting effect with the same parameters as in Fig. 3. The roughly parabolic contours seen on the right of Figs. 2, 4, and weakly in Fig. 5 are truncated by an arc curving between the points $(\phi,\theta)=90°, +\theta_c$; $\sim 60°$, 0; $90°, -\theta_c$. This discontinuity is the line, calculated by invariants[21] as

$$\cos\ \theta_c = \cos\ \theta \sin\ \phi, \tag{25}$$

which defines all rays for which the critical angle is just reached as the ray azimuth reaches $\pi/2$. Since $\theta_c = 30°$ in this case the value of ϕ for $\theta = 0$ is indeed 60°. To the left of this line upslope rays never hit the critical angle and continue indefinitely bending round to deep water. To the right they are always truncated and therefore weaker.

An interesting question to ask is, when can we ignore the 3D effects and still get the same directionality? The simple answer is that the exponent in Eq. (12) can never be the same evaluated by the two methods in the direction slightly up from across-slope. However we can make the overall effect negligible if we guarantee the exponent to be greater than one. This can be done by ensuring that:

$$\alpha/2|\epsilon_0| \gg 1. \tag{26}$$

In fact, Figs. 2 and 4, which are similar for most angles (but not identical), have $\alpha/2|\epsilon_0| = 11.5$, whereas Figs. 3 and 5, which differ significantly have $\alpha/2|\epsilon_0| = 1.15$, thus reinforcing this rule. In short, we can neglect 3D effects if losses are high or slopes are low.

IV. SLOPING BOTTOM WITH RANGE-INDEPENDENT REFRACTION

In a range-independent refracting environment one can only receive noise contributions (other than by direct paths) by surface duct (SD) or by surface/bottom reflections (SB); other paths never reach the surface.[19] We now consider a ''range-dependent environment,'' defined as variable bathymetry but fixed sound speed profile. Not only do we have possible angle changes with range, but we can also have conversion from one type of path to another. Thus we may have contributions from sound channel (SC) and bottom duct (BD) paths as seen in Fig. 1(b) and (c). This phenomenon is well known,[13] and attempts have been made to understand the uniformity of the received angle distribution under these conditions.[17] Rather than address the general problem we confine our attention to the simpler one of a uniform bottom slope with a range-independent refracting medium. We consider three cases: downward refraction, sound channel, and upward refraction. Here the conversion rules are quite straightforward.

A. Downward refraction with uniform bottom slope (BD/SB)

Returning to the elevation angle dependence of the intensity (D) in the $N\times 2D$ representation we now recalculate Eqs. (17) and (18) for various cases. In passing we shall see a simple process for providing a ''flat'' angle response in what would otherwise be a ''noise notch.'' In Fig. 1(b) if we set $\theta_r = 0$, the horizontal ray contributes no noise from the downslope direction, but is converted to SB some way up the slope. This is the furthest upslope that any of this BD family of rays can be converted. The steepest BD ray at the receiver has $\theta_r = \theta_0$, such that it just grazes the surface [where $c(z) = c_s$] in its first cycle

$$\cos\ \theta_0 = \frac{c_r}{c_s}. \tag{27}$$

It also contributes nothing from downslope but is immediately converted to SB upslope. This angle range (together with the downgoing ray equivalent, i.e., $-\theta_0$ to $+\theta_0$) is exactly what would have constituted a noise notch if the bottom slope had been zero. Inside this range there is always ultimately a SB contribution similar to Eq. (21) for a surface angle range zero up to the surface corrected critical angle θ_{sc}. To evaluate this with approximate refraction effects we return to Eq. (10), including surface and bottom losses

$$S = \int \sin\ \theta_s e^{-\{\int \alpha_s \sin\ \theta_s\, dj + \int \alpha_b \sin\ \theta_b\, dj\}}\, dn. \tag{28}$$

If we make the reasonably good approximation in a downward refracting environment that the surface angle increments by fixed amounts per bounce then

$$dj = d\theta_s/2\epsilon, \tag{29}$$

and the first integral in the exponent becomes

$$(\alpha_s/2|\epsilon|)(\cos\ \theta_{s1} - \cos\ \theta_s),$$

θ_{s1} being the initial value of θ_s.

In the second integral we make the poorer (and slightly inconsistent) approximation that $dj = d\theta_b/2\epsilon$, resulting in $(\alpha_b/2|\epsilon|)(\cos\ \theta_{b1} - \cos\ \theta_b)$. However we can write this as

$$(\alpha_b/2|\epsilon|)(c_{b1} \cos\ \theta_{s1} - c_b \cos\ \theta_s)/c_s,$$

where c_{b1}, c_b are the local bottom sound speeds. By comparison with the above approximations, the assumption that $c_{b1} \simeq c_b \simeq c_s$ is good so, in effect, we can use the terms $(\cos\ \theta_{s1} - \cos\ \theta_s)$ and $(\cos\ \theta_{b1} - \cos\ \theta_b)$ interchangeably. We can then solve the main integral by invoking Eq. (29) and the result depends on the joint boundary loss $\alpha = \alpha_s + \alpha_b$.

The final contribution from upslope including the SB and the BD sections for $|\theta_r| < \theta_0$ is

$$S_{\rm up}(\phi,\theta_r) = I_{\rm BD}(\phi,\theta_r)\, \frac{1}{\alpha}\, (1 - e^{-(\alpha/2|\epsilon|)(1-\cos\ \theta_{sc})}), \tag{30}$$

where $I_{\rm BD}(\theta_r)$ is the BD attenuation out to the transition. Note that the SB contribution is *independent of angle at the receiver*.

For a linear profile of gradient c' we have

$$I_{\rm BD}(\phi,\theta_r) = \exp(-\alpha_b \sin\ \theta_b r_t/r_c), \tag{31}$$

where r_t is the transition range and r_c is the cycle distance. These are easily shown to be

$$r_t = \frac{c_r}{|c'||\epsilon|}\, (\sec\ \theta_0 - \sec\ \theta_r), \tag{32}$$

$$r_c = \frac{2c_b}{|c'|} \tan\ \theta_b. \tag{33}$$

This results in

$$I_{\mathrm{BD}}(\phi,\theta_r)=\exp\left(-\frac{\alpha_b}{2|\epsilon|}\left(\frac{c_s}{c_r}\cos\,\theta_r-1\right)\right), \tag{34}$$

which ensures continuity when $|\theta_r|=+\theta_0$ by tending to unity through Eq. (27).

If bottom losses are significant then we expect $I_{\mathrm{BD}}(\phi,\theta_r)$ and consequently $S(\phi,\theta_r)$ to dip in the middle. This behavior is shown at the center right of Fig. 6. In the equivalent downslope angle there is still a noise notch.

Outside this angle range behaviour is almost identical to the isovelocity case [Eqs. (20) and (21)] except that we use the angle at the surface θ_s, where

$$\cos\,\theta_s=\frac{c_s}{c_r}\cos\,\theta_r. \tag{35}$$

From Eq. (27) $\theta_s=0$ at the surface when $\theta_r=\theta_0$.

In the downslope direction we now have

$$S_{\mathrm{down}}(\phi,\theta_r)=\frac{1}{\alpha}\,(1-e^{-(\alpha/2|\epsilon|)(1-\cos\,\theta_s)}), \tag{36}$$

and in the upslope direction we have

$$S_{\mathrm{up}}(\phi,\theta_r)=\frac{1}{\alpha}\,(1-e^{-(\alpha/2|\epsilon|)(\cos\,\theta_s-\cos\,\theta_{sc})}). \tag{37}$$

B. Sound channel with uniform bottom slope (SC/BD/SB)

In Fig. 1(c) we can imagine two limiting rays, one setting off horizontally, the other grazing the surface. If we assume the same velocity at surface and bottom with a minimum in between, then again this angle range $|\theta_r|<\theta_0$ corresponds to a noise notch in the zero slope case. Each ray has potentially a transition from SC to BD, then BD to SB.

Inside this angle range we find no contribution from downslope, and in the upslope direction we have Eq. (30) multiplied by $I_{\mathrm{SC}}(\phi,\theta_r)$, i.e., the attenuation in the sound channel up to the first transition. Now, a subtle distinction between this and the downward refraction case is that it is possible for this attenuation to be low (i.e., volume absorption only) ($I_{\mathrm{SC}}\rightarrow 1$) and simultaneously for the BD section to be short lived ($I_{\mathrm{BD}}\rightarrow 1$); this can happen if the bottom is steep enough for its slope to be comparable with the sound channel ray angle. We then have

$$S_{\mathrm{up}}(\phi,\theta_r)=\frac{1}{\alpha}\,(1-e^{-(\alpha/2|\epsilon|)(1-\cos\,\theta_{sc})}), \tag{38}$$

which is independent of θ_r and explains the flat angle response or ''pedestal'' sought by Dashen and Munk.[17] This is clearly seen at the center right of Fig. 7.

Outside this angle range behavior is as in the bottom duct provided either the SSP is isovelocity below the sound channel, or water depths are held constant on the ''deep'' side of the receiver. If, in contrast, sound speed is allowed to increase with depth there will be an extra SD contribution similar to that seen in the next section.

One might object that, although there is no angle dependence for $|\theta_r|<\theta_0$, a finite I_{BD} would lead to a low barrier at the edge of the flat section. In practice all higher angle contributions can be arbitrarily reduced by invoking stronger noise sources near the shelf region than in the open ocean.

C. Upward refraction with uniform bottom slope (SD/SB)

Figure 1(d) shows SD rays in transition to SB. Again we can imagine two limiting rays, one horizontal, and one just grazing the bottom. Thus there is an angle range $|\theta_r|<\theta_0$ within which SD contributions arrive at the receiver, and indeed there would be contributions even for zero bottom slope. Within this angle range the contribution from down slope is exactly as in Harrison,[19]

$$S_{\mathrm{down}}(\phi,\theta_r)=\frac{1}{\alpha_s}. \tag{39}$$

In the upslope direction we now have two contributions: one similar to Eq. (30) from the attenuated SB returns, and one from the truncated SD itself[19] which turns out to contain a term F that is identical to I_{SD}:

$$S_{\mathrm{up}}(\phi,\theta_r)=I_{\mathrm{SD}}(\phi,\theta_r)\times\frac{1}{\alpha}\,(1-e^{-(\alpha/2|\epsilon|)(1-\cos\,\theta_{sc})}) + \frac{1}{\alpha_s}\,(1-F), \tag{40}$$

and by reasoning similar to that in Eqs. (32)–(34), we have

$$I_{\mathrm{SD}}(\phi,\theta_r)=F=\exp\left(-\frac{\alpha_s}{2|\epsilon|}\left(\frac{c_b}{c_r}\cos\,\theta_r-1\right)\right) \tag{41}$$

in terms of the bottom sound speed c_b below the receiver.

Outside this angle range in the upslope direction behavior is exactly as in Eq. (27) allowing for inverting the SSP:

$$S_{\mathrm{up}}(\phi,\theta_r)=\frac{1}{\alpha}\,(1-e^{-(\alpha/2|\epsilon|)(\cos\,\theta_b-\cos\,\theta_c)}). \tag{42}$$

In the downslope direction there is one term similar to Eq. (36) and another from the distant converted SD:[19]

$$S_{\mathrm{down}}(\phi,\theta_r)=\frac{1}{\alpha}\,(1-e^{-(\alpha/2|\epsilon|)(1-\cos\,\theta_b)}) + I_{\mathrm{SB}}(\phi,\theta_r)\,\frac{1}{\alpha_s}, \tag{43}$$

where

$$I_{\mathrm{SB}}(\phi,\theta_r)=e^{-(\alpha/2|\epsilon|)(1-\cos\,\theta_b)}, \tag{44}$$

and θ_b is the local bottom angle corresponding to θ_r, i.e.,

$$\cos\,\theta_b=\frac{c_b}{c_r}\cos\,\theta_r. \tag{45}$$

This behavior is shown in Fig. 8.

V. CONCLUSIONS

This paper has extended an earlier treatment of noise directionality to range-dependent environments, paying particular attention to the effects of bathymetry. Using a very simple treatment, closed-form formulae have been derived for the dependence on azimuth ϕ and elevation angle θ_r at

TABLE I. Summary of equation numbers for various regimes. Limiting angle θ_0 is defined by Eqs. (27), (35), or (45), as appropriate.

Environment	Angle	Downslope	Upslope
Isovelocity (SB)	All	20	21
Downward	$<\theta_0$	⋯	30+34
refraction (BD)	$>\theta_0$	36	37
Sound	$<\theta_0$	⋯	30+34/38
channel (SC)	$>\theta_0$	36	37
Surface	$<\theta_0$	39	40+41
duct (SD)	$>\theta_0$	43+44	42

the receiver in a wedge shaped ocean. A summary of equation numbers for the various regimes is given in Table I.

In the isovelocity case, results for the $N\times 2D$ azimuthal slice approach (Fig. 2) were contrasted with a full 3D approach (Fig. 4), including horizontal bending. It was shown that significant effects are possible particularly when losses are low or slopes are significant (Figs. 3 and 5). These are also the conditions under which the directionality is most acute.

Three other cases with fixed sound speed profile and sloping bottom were considered. First, in the downward refraction case there is an angle region that would otherwise be occupied by the noise notch. This is filled on the upslope side by surface/bottom reflections that convert to bottom duct on their way to the receiver. There is a slight dip in the centre of this angle range (Fig. 6) because of bottom losses. In the downslope direction the noise notch is unchanged. The second case, a sound channel, has similar features (Fig. 7), except that the central region may be flat in elevation angle as has been found experimentally. Finally, the upward refracting case differs in that there is no noise notch, and in the upslope and downslope directions there are hybrid contributions partly from surface duct and partly from surface/bottom reflections (Fig. 8).

ACKNOWLEDGMENTS

This work was supported by BAeSEMA Ltd. Related work on the general purpose ambient noise model CANARY has been funded by the U.K. Defence Research Agency. I am grateful to A. Cowley of the DRA for bringing the subject of ambient noise to my attention, and Dr. R. Brind of DRA who has recently encouraged development of CANARY for range- and azimuth-dependent environments.

[1] B. F. Cron and C. H. Sherman, ''Spatial correlation functions for various noise models,'' J. Acoust. Soc. Am. **34**, 1732–1736 (1962).
[2] R. A. Wagstaff, ''RANDI: Research Ambient Noise Directionality Mode,'' Naval Undersea Center, TP 349, April 1973.
[3] J. E. Breeding, L. Pflug, M. Bradley, M. Hebert, and M. Wooten, ''RANDI 3.1 User's Guide,'' Naval Research Laboratory Technical Memorandum NRL/MR/7176-94-7552, Stennis Space Center, MS, 1994.
[4] R. W. Bannister, D. J. Kewley, and A. S. Burgess, ''Directional Underwater Noise Estimates—the DUNES model,'' Weapons Systems Research Laboratory, Tech Note WSRL-TN-34/89, 1989.
[5] W. Renner, ''Ambient Noise directional estimation system (ANDES) II User's Guide,'' Science Applications International Corporation Technical Memorandum SAIC-88/1567, 1988.
[6] C. H. Harrison and A. Cowley, ''CANARY: A model of ambient noise and coherence,'' in Third European Conference on Underwater Acoustics, edited by J. S. Papadakis (1996).
[7] C. H. Harrison and A. Cowley, ''A model of noise response for baffled hydrophone arrays,'' in *Arrays and Beamforming in Sonar*, edited by P. Dobbins (IOA, London, 1996), IOA Proceedings, Vol. 18, Pt. 5.
[8] R. Hamson, ''The modeling of ambient noise due to shipping and wind sources in complex environments,'' Appl. Acoust. **51**(3), 251–287 (1997).
[9] J. S. Perkins, W. A. Kuperman, F. Ingenito, L. T. Fialkowski, and J. Glattetre, ''Modeling ambient noise in three-dimensional ocean environments,'' J. Acoust. Soc. Am. **93**, 739–752 (1993).
[10] W. A. Kuperman, M. B. Porter, J. S. Perkins, and R. B. Evans, ''Rapid computation of acoustic fields in three-dimensional ocean environments,'' J. Acoust. Soc. Am. **89**, 125–133 (1991).
[11] W. A. Kuperman and F. Ingenito, ''Spatial correlation of surface generated noise in a stratified ocean,'' J. Acoust. Soc. Am. **67**, 1988–1996 (1980).
[12] W. M. Carey and R. A. Wagstaff, ''Low frequency noise fields,'' J. Acoust. Soc. Am. **80**, 1523–1526 (1986).
[13] R. A. Wagstaff, ''Low frequency ambient noise in the deep sound channel—The missing component,'' J. Acoust. Soc. Am. **69**, 1090–1014 (1981).
[14] V. C. Anderson, ''Variation of the vertical directionality of noise with depth in the North Pacific,'' J. Acoust. Soc. Am. **66**, 1446–1452 (1979).
[15] G. R. Fox, ''Ambient noise directionality measurements,'' J. Acoust. Soc. Am. **36**, 1541–1546 (1964).
[16] G. B. Morris, ''Preliminary results on sea mount and continental slope reflection enhancement of shipping noise,'' SIO Ref 75-34, Scripps Institution of Oceanography, La Jolla, CA, 1975.
[17] R. Dashen and W. Munk, ''Three models of global ocean noise,'' J. Acoust. Soc. Am. **76**, 540–554 (1984).
[18] B. J. Sotirin and W. S. Hodgkiss, ''Fine-scale measurements of the vertical ambient noise field,'' J. Acoust. Soc. Am. **87**, 2052–2063 (1990).
[19] C. H. Harrison, ''Formulas for ambient noise level and coherence,'' J. Acoust. Soc. Am. **99**, 2055–2066 (1996).
[20] D. E. Weston, ''Guided propagation in a slowly varying medium,'' Proc. Phys. Soc. London **73**, 365–384 (1959).
[21] C. H. Harrison, ''Three-dimensional ray paths in basins, troughs and near seamounts by use of ray invariants,'' J. Acoust. Soc. Am. **62**, 1382–1388 (1977).
[22] D. E. Weston, ''Propagation in water with uniform sound velocity but variable depth lossy bottom,'' J. Sound Vib. **47**, 473–483 (1976).
[23] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970), Sec. 4.3.149.

Acoustic reverberation from a laboratory model of a shelf break

Saimu Li and C. S. Clay
Geophysical and Polar Research Center, University of Wisconsin—Madison, 1215 W. Dayton Street, Madison, Wisconsin 53706

(Received 17 July 1995; accepted 29 July 1997)

Continental boundaries have a shallow sloping bottom that changes to a steeper slope. Our laboratory acoustic models (simulations) have (1) a 11° slope that changes to a 50.5° slope and (2) a 35° slope change to a 59° slope. The surface and bottom of the acoustic model are dry wall construction board. The source is a spark and the receiver is a small microphone. The Biot–Tolstoy exact time-domain solution was used. [I. Tolstoy and C. S. Clay, *Ocean Acoustics* (American Institute of Physics, New York, 1987), 2nd ed., Appendix 5, and C. Feuillade and C. S. Clay, J. Acoust. Soc. Am. **96**, 501–514 (1994)]. The image-reflection and diffraction components depend on source and receiver locations. In model (1), source and receiver were within the 11° wedge and gave reverberations that were mainly from the 11° wedge. Source and receiver beyond the shelf break gave reverberations that were mainly from the 50° part of the model. In model (2), the source was up slope and receiver was near the break to observe diffractions. The finite number of arrivals (images) obeyed the Biot–Tolstoy rule. Experimental signals matched the theory in amplitudes and arrival times. Most reverberations were caused by multiple reflections in the wedge waveguide. The amplitudes of the image-reflection paths were an order of magnitude larger than the diffraction arrivals. © *1997 Acoustical Society of America.* [S0001-4966(97)04211-2]

PACS numbers: 43.30.Zk, 43.20.Dk, 43.20.Mv, 43.20.Px [JHM]

LIST OF SYMBOLS

b	$\equiv c\sqrt{2\tau_0/(rr_0)}$
c	sound speed
k,n	integers
P_0	source amplitude at R_0
$p(t)$	acoustic pressure. Subscripts give name, R for reflection, D for diffraction
r_0	source range
r	receiver range
R_0	reference range, 1 m
$R_k(\pm)$	total path distances for image paths
t	time s. Subscripts give particular times
T_b	averaging time for first time step
t_0	time step increment
x,y,z	coordinates in x,y,z rectangular system
δ	a small loss term for an imperfect wedge
$\delta_f(t)$	finite delta function of duration Δt
Δt	duration of finite impulse
θ_w	wedge angle
θ_0	source angle
θ	receiver angle
v	$=\pi/\theta_w$
η	function of time, wedge angle, source and receiver positions

INTRODUCTION

We have had excellent results in comparing the diffractions and reflected experimental acoustic data with the Biot–Tolstoy theory.[1–10] Chambers and Berthelot also show excellent agreement with theory.[11] We choose the shelf break type of structure shown in Fig. 1. This kind of structure is a typical reverberation problem with diffractions and multiple reflections. This is clearly a typical reverberation problem. These kinds of structure are common. An elegant reverberation study has been published by Preston and Kinney.[15] We expect to find considerable reverberation from multiple reflections from the wedge facets and the diffractions from the wedges.

The surface is O to C. The shelf is from O to A. The break is at A and the continental slope A to B. D is the intercept of B–A at the surface. S and R stand for source and receiver. The low-pass filter was 30 kHz and the high-pass filter was 2 kHz. The analog signals were sampled at 100 kHz and stored in the computer memory. Fifty signal transmissions were stacked; the mean value was deleted and then stored as a data file. Here the source S and receiver R are shown at positions off shore from the shelf break at A. The transmission measurements were made in air, $c = 346.6$ m/s.

In previous studies[6,8] the wedge waveguide was a single wedge. Here, our laboratory models simulate the structure of a continental boundaries such as the east coast of United States. It has a shallow sloping bottom (0°–0.5°) that changes to a steeper slope (6°–11°) at the continental slope. Our laboratory acoustic model (Fig. 1) greatly exaggerates the actual slopes. The model of the shelf and break has a 10.8° slope angle C–O–A for the shelf. The slope changes to approximately 50° at the shelf break A. The surface and bottom of the acoustic model are made of dry wall construction board. The source is a spark (S) and the receiver (R) is a small electret microphone (2 mm×2 mm). This is an extension of our earlier research.

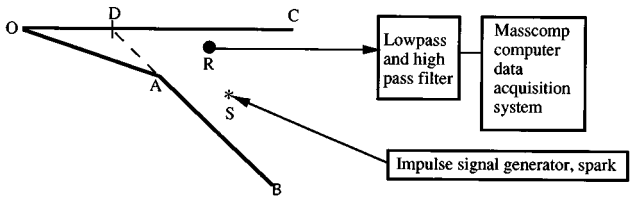


FIG. 1. Shelf and slope geometry and the experimental apparatus. The surface is O to C. The shelf is from O to A. The break is at A and the continental slope A to B. D is the intercept of B-A at the surface. S and R stand for source and receiver. The low-pass filter was 30 kHz and the high-pass filter was 2 kHz. The analog signals were sampled at 100 kHz and stored in the computer memory. Fifty signal transmissions were stacked; the mean value was deleted and then stored as a data file. Here the source S and receiver R are shown at positions off shore from the shelf break at A. The transmission measurements were made in air, $c=346.6$ m/s.

I. BIOT–TOLSTOY WEDGE FORMULAE

The Biot–Tolstoy (B–T) exact time-domain solution was used as basis for the numerical calculations.[4] Feuillade and Clay give a detailed development of the image part of the exact solution. In addition to the original Biot–Tolstoy paper, there are several expositions that give ''simplifying approximations'' for special conditions and Fourier transformations.[13,14] Buckingham's frequency domain formulation is not used because it takes all modes and Fourier transformation to time domain to compute the sound pressure for the arrivals.[15]

We make minor changes in the exact B–T expressions for our numerical computations by including a small damping constant δ. Physically, the damping constant represents an imperfect intersection of the two planes of the wedge. Examples of wedge constructions are shown in Fig. 2. The light and dark shading represent two media. In these models, the sound speeds are the same and the densities are different. For rigid boundaries, the density becomes infinite and for free boundaries, the density becomes zero.

The exact solution for an impulsive source near a wedge has two components: (1) The reflections from the planes that form the wedge, and (2) the diffraction from the line of intersection of the planes. Biot and Tolstoy showed that the reflected components reduce to image constructions. The diffracted component is an impulsive pressure that decays rapidly. The pressure signal is the sum of the direct and reflected components $p_R(t)$, as well as a diffracted component $p_D(t)$

$$p(t)=p_R(t)+p_D(t). \tag{1}$$

Many acoustic experiments use explosive sources, i.e., sparks, that give a delta function of pressure. In our discussion we use the finite/discrete $\delta_f(t)$, unit s^{-1}. The sound pressure has a peak pressure P_0 referred to R_0, the duration of the peak is Δt, and the impulse is $P_0\Delta t$

$$\delta_f(t)=1/\Delta t \quad \text{for} \quad 0\leqslant t\leqslant \Delta t \quad \text{and}$$

$$=0 \quad \text{otherwise.}$$

A. Image region

The sound pressure in the image region is

$$p(t)=P_0\Delta t R_0 \sum_k \frac{\delta_f[t-R_k(\pm)/c]}{R_k(\pm)}, \tag{2}$$

where the $R_k(\pm)$ are the combinations that are given by

$$R_k(\pm)=[r_0^2+r^2+y^2-2r_0 r\ \cos(2k\theta_w\pm\theta\pm\theta_0)]^{1/2}. \tag{3}$$

Each of the image contributions is represents the energy for a normal coordinate. In region 2 there are a discrete number of normal coordinates. The maximum value of k or $k_{\max}$ is given by the condition

$$0<|2k\theta_w\pm\theta_0\pm\theta|\leqslant\pi. \tag{4}$$

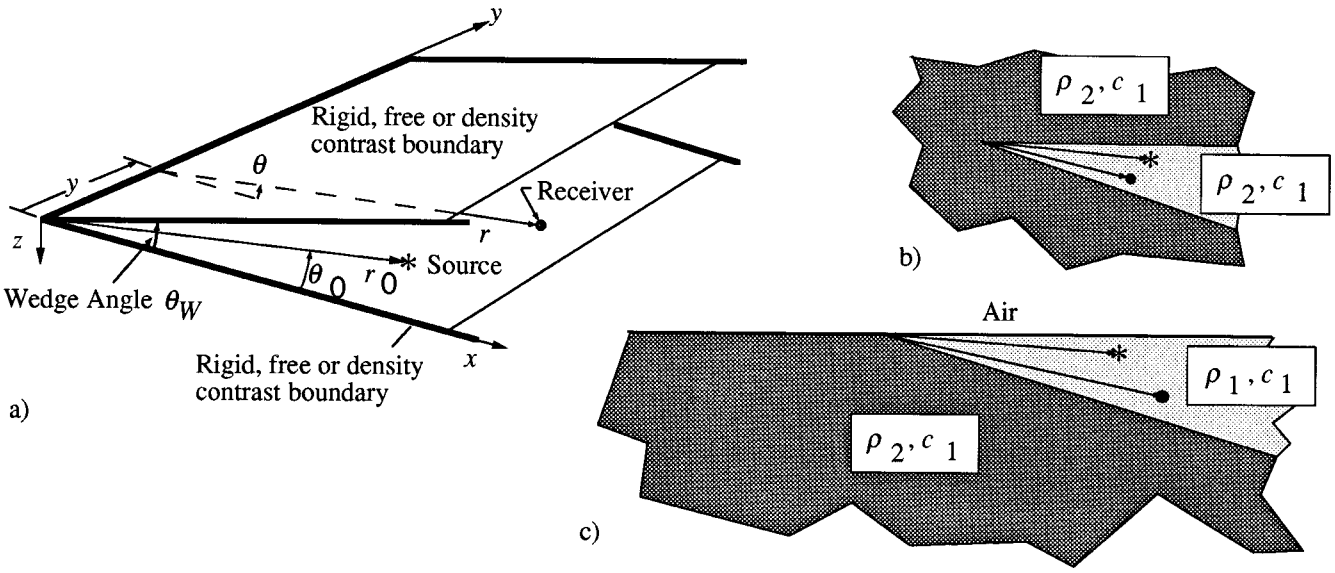


FIG. 2. Wedge constructions and density contrast wedges. The light and dark shading represent two media. In these models, the sound speeds are the same and the densities are different. For rigid boundaries, the density becomes infinite, and for free boundaries, the density becomes zero.

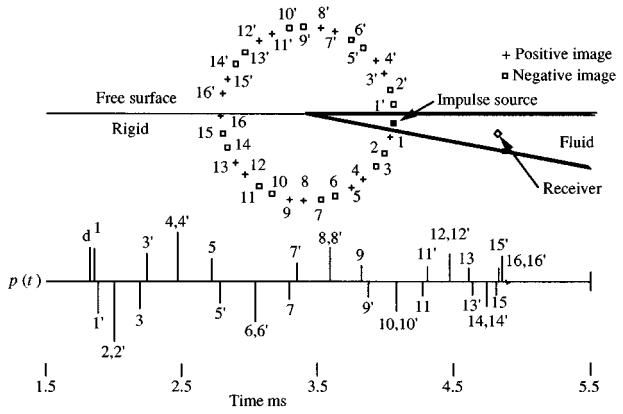


FIG. 3. Example of an image constructions and signals. The images below the fluid wedge are numbered 1,2, etc. Images above the free surface are numbered $1',2'$, etc. The source to receiver direct travel path is labeled ''d.'' The path from the image 1 to the receiver is labeled 1 where the R is dropped because all paths go to the receiver. Time is measured after the initiation of the impulse. The parameters are: wedge angle $\theta_w = 10.8°$, $\theta_0 = \theta = 6°$, $R_0 = 0.53$ m, $R = 1.16$ m, $y = 0$, $\mathscr{R}_{10} = -1$, $\mathscr{R}_{12} = 1$, and in the fluid $c = 346.6$ m/s. (a) Free-rigid boundaries positive image pairs and negative image pairs. (b) Arrivals due to multiple reflections in the wedge, i.e., the arrivals from the images. ''d'' is the direct arrival. Travel times are relative to $t = 0$ at the source. Arrivals 2 and $2'$, 4 and $4'$, etc. arrive at the same times and their pressures add.

This condition gives a finite number of images. The pressure disturbance from the last image arrives before the arrival of the diffraction from the line of intersecting planes. We show an image construction in Fig. 3. For illustration, one plane of the wedge is free and the other is rigid and this gives alternating signs for the images. The images are reflections of the source in wedge planes. Possible image locations are on the circle containing the source and at the angles given by $2k\theta_w \pm \theta_0$. Here k ranges from $-k_{\max}$ to $+k_{\max}$.

The sound pressures of the arrivals $p(t)$ are also shown in Fig. 3. The number of the image and the corresponding arrival are identified. In this illustration, the choice $\theta_0 = \theta$, causes some pairs of the arrivals to have the same travel times and the sum of the sound pressures is twice the pressure amplitude of the arrival from a single image.

The images below the fluid wedge are numbered 1, 2, etc. Images above the free surface are numbered $1',2'$, etc. The source-to-receiver direct travel path is labeled ''d.'' The path from the image 1 is labeled 1, where the R is dropped because all paths go to the receiver. Time is measured after the initiation of the impulse. The parameters are: wedge angle $\theta_w = 10.8°$, $\theta_0 = \theta = 6°$, $R_0 = 0.53$ m, $R = 1.16$ m, $y = 0$, $\mathscr{R}_{10} = -1$, $\mathscr{R}_{12} = 1$, and in the fluid $c = 346.6$ m/s. (a) Free-rigid boundaries positive image pairs and negative image pairs. (b) Arrivals due to multiple reflections in the wedge, i.e., the arrivals from the images. ''d'' is the direct arrival. Travel times are relative to $t = 0$ at the source. Arrivals 2 and $2'$, 4 and $4'$, etc. arrive at the same times and their pressures add.

The choice of a receiver angle θ requires that satisfaction (4) be tested to determine if one can receive an arrival from that image. The positive values of k use $(2k\theta_w \pm \theta_0 - \theta)$ and the negative values of k use $(2k\theta_w \pm \theta_0 + \theta)$:

$$t_k(\pm) = \frac{R_k(\pm)}{c}. \tag{5}$$

In numerical computations and at a sequence of time steps jt_0, the jth time step is

$$j = \mathrm{INT}[t_k(\pm)/t_0], \tag{6}$$

where integer operation is ''INT,'' in Basic or ''round'' in MATLAB. The small fluctuation of the arrival time within a time step is acceptable.

The raypaths from the images to the receivers in the density contrast model are straight because the sound speeds are the same in all regions. The reflection coefficients and transmission coefficients are constant for all angles of incidence

B. Diffraction arrival

The derivation of the diffraction arrival is in Refs. 1–4, and we give the the sound pressure $p_D(t)$ due to an impulsive pressure source. The diffraction arrival exists for $t > \tau_0$, where

$$\tau_0 = \frac{1}{c}\left[(r + r_0)^2 + y^2\right]^{1/2}, \tag{7}$$

where τ_0 is the minimum time for a disturbance to go from the source to the line of intersection of the planes and to the receiver:

$$p_D(t) = \frac{-P_0 \Delta t R_0 c}{2\theta_w r r_0}\left[D(\eta, +-) + D(\eta, -+) + D(\eta, ++) + D(\eta, --)\right], \tag{8}$$

where

$$\eta = \log[U + (U^2 - 1)^{1/2}], \tag{9}$$

$$U \equiv \frac{c^2 t^2 - (r^2 + r_0^2 + y^2)}{2 r_0 r}, \tag{10}$$

and η goes to 0 as t tends to τ_0. The diffraction term, $D(\eta, +-)$, is

$$D(\eta, +-) \equiv \frac{\sin\ \nu(\pi + \theta - \theta_0)}{\sinh(\eta + \delta)\{\cosh[\nu(\eta + \delta)] \pm \cos\ \nu(\pi + \theta - \theta_0)\}},$$
$$\nu = \pi/\theta_w, \tag{11}$$

where we added δ to form $(\eta + \delta)$ and to include the imperfection of the wedge. Now the $D(\eta, +-)$ terms are finite as t tends to τ_0. It is convenient to measure the dependence of the diffraction component $p_D(t)$ from τ_0

$$\tau = t - \tau_0, \tag{12}$$

and U becomes

$$U = 1 + \frac{c^2 \tau^2}{2 r_0 r} + \frac{c^2 \tau \tau_0}{r_0 r}. \tag{13}$$

Diffraction expressions for different boundaries are in Ref. 4.

C. Numerical computations

In the time expression $t=\tau_0+\tau$, the time τ after the beginning of the diffraction arrival τ_0 must be accurately computed. One can compute accurate values of τ by using very small sampling steps and measuring all distances in the sampling steps ct_0. A simpler robust method uses a secondary time reference to compute τ as follows:

$$j_{\text{tau}}=\text{INT}(\tau/t_0); \tag{14}$$

and the secondary reference is

$$j_{\text{dif}}=\text{INT}(\tau_0/t_0). \tag{15}$$

The nth time step for the sound pressure is

$$n=\text{INT}(t/t_0); \tag{16}$$

$$n=j_{\text{dif}}+j_{\text{tau}}. \tag{17}$$

The reference array element for τ_0 is j_{dif} and at $\tau=0$, $j_{\text{tau}}=0$. The diffraction terms are computed for $n=0,1,...$, $j_{\text{dif}}+0,+1...$ etc.

II. IMAGES, DIFFRACTIONS, AND MORE IMAGES

The geometry for the first simulation is shown in Fig. 4(a). There are three wedges: (1) The ''beach'' shelf wedge is A–O–C (2) The shelf break at *A* is the apex of the wedge O–A–B. (3) The planes B–A and D–C form the incomplete wedge, B–D–C, where D is the projection of B–A on A–C. Both the source and receiver are ''off shore'' from the wedge break at A. Following Ref. 4, Fig. 4(b) shows a kaleidoscopic construction of the images of the source around the apex (or intersection of the wedge planes) and the images of the surfaces. Then, we connect the source images to the receiver by straight lines. If a given line passes through all the surfaces, then this raypath is allowed. *If the line misses even one surface, the ray is not allowed*. For example, the raypaths from the images 1, 2, 3, and 7 are allowed. The raypaths from images 4, 5, and 6 are not allowed. Even small changes of geometry can cause the raypath from 7 to the receiver to fail because the raypath is on the edge of the wedge plane. [Again, the receiver is the end of the raypaths and the path 1 to *R* is designated as 1.]

Figure 4(c) shows the kaleidoscopic construction for diffraction arrivals from *A* and its image A′.[4] For this geometry, the travel times for the reflection path 7 and the diffraction path 3A are nearly the same. The pressure signal at the receiver is shown in Fig. 4(d). The raypaths for the arrivals are labeled. The results are sensitive to the choices of θ_W, θ_0, and θ. If $\theta_0+\theta=\theta_W$, then the arrivals $1-R$ and $1'-R$ arrive at the same times and the pressures add. If $\theta_0=\theta$, then other arrivals have the same travel times and the pressures add.

III. SIMULATION 1

The waveguide used in the experiment is shown in Figs. 1 and 4. All of the experiments used the impulsive transmissions from a spark source to a small microphone in air. A typical transmission in free air was shown in Fig. 5. Later, this signal was used as the reference source function for con-

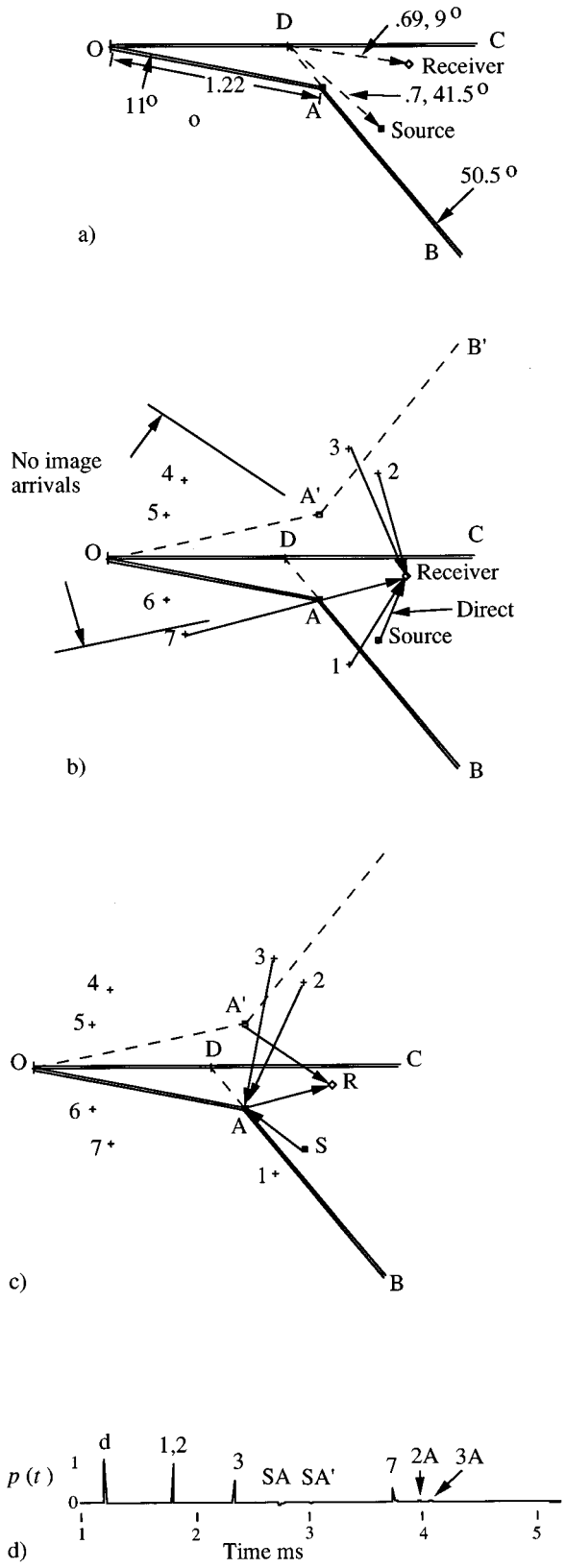


FIG. 4. Geometry for simulation 1. The source and receiver are beyond the shelf break at A. Here the primary wedge for the image constructions is B–D–C. (a) The model geometry, source and receiver positions for simulation 1. (b) Image constructions for the source. Here, the images are labeled 1 to 7. The active images are 1, 2, 3, and 7. ''Raypaths'' from images 4, 5, and 6 do not pass through the reflecting facet A–B and A′–B′. (c) The diffractions come from A and its image A′. Diffractions come from the source and its image. (d) Simulation of the travel times and amplitudes. The direct travel time is 1.18 ms.

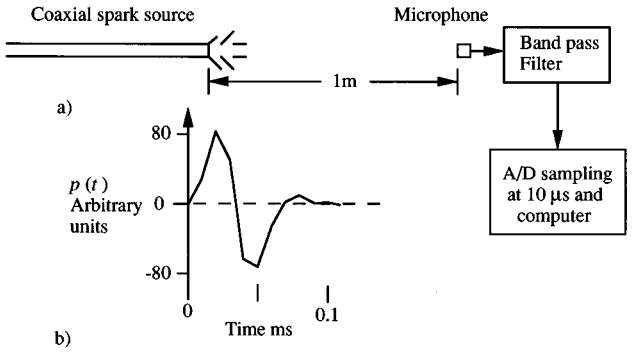


FIG. 5. ''Free-air'' transmission of sound pressure from the spark. The spark signal was band pass filtered 2–30 kHz and then sampled at 100 KHz. The source-to-microphone travel time is suppressed. $c=346.6$ m/s. The amplifier gains were constant so that the transmission was a calibration of the source–receiver system in arbitrary units. Although not shown here, our coaxial spark excites a (tube) wave that travels backward along the coaxial tube, reflects at the end, and then travels forward along the tube. This wave then radiates as a pressure signal from the spark end of the tube. The length of the tube, 0.9 m, limits the duration of an experiment to about 5 ms. The construction of the coaxial spark described in the Appendix.

volutions with theoretical impulse response for all of the comparisons. The spark signal was band pass filtered 2 to 30 kHz and then sampled at 100 kHz. The source-to-microphone travel time is suppressed. $c=346.6$ m/s. The amplifier gains were constant so that the transmission was a calibration of the source–receiver system in arbitrary units. The coaxial spark described in the Appendix.

The convolution of the spark signal Fig. 5 and the impulse response, Fig. 4(d), is shown in Fig. 6(a). The arrivals are identified with the nomenclature in Figs. 3 and 4(d). The experimental arrangement and impulse response are shown in Fig. 4. The identifications of arrivals are modified from Fig. 4(d). Travel times are measured following initiation of the spark. The direct arrival is at 1.18 ms.

The reflected components or image arrivals much larger than the background noise. They have the same amplitudes, arrival times, and wave forms as the theory. The strongest diffraction arrival is $2'$–A–R. It is likely that $2'$–A–R and $2'$–A$'$–R arrive at nearly the same times because the small

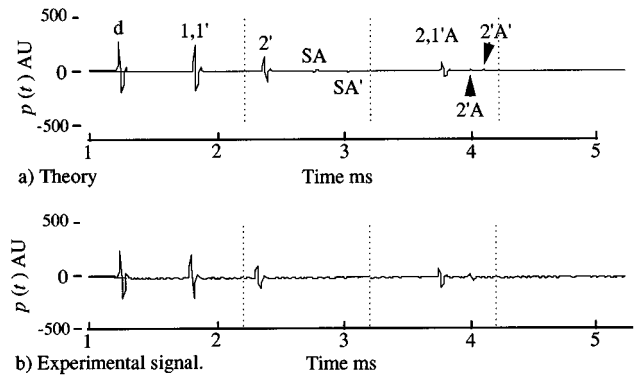


FIG. 6. Simulation 1: Comparison of theory and experiment. The geometry is in Fig. 4(a) and the impulse response is Fig. 4(d). Change the nomenclature as follows: 1=1; 2=1′; 3=2′; 7=2,1′A; 3A=2′A. Travel times are measured following initiation of the spark. The direct arrival is at 1.18 ms.

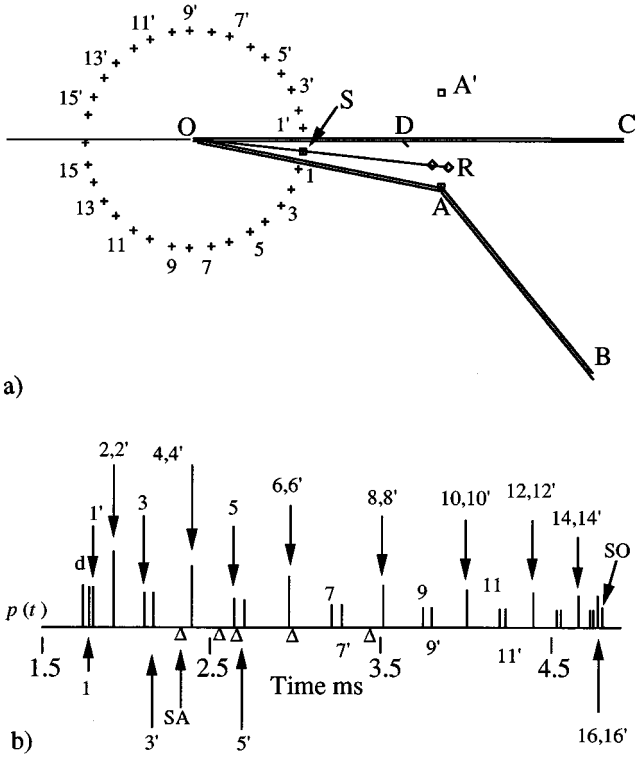


FIG. 7. Simulation 2: Image construction and impulse response. The source impulse is at $t=0$. Wedge angle $\theta_w=10.8°$. Source and receiver angles are $\theta=\theta_0=6°$. The source range from the apex O is OS=0.53 m. The receiver ranges are OR=1.16 m to 1.24 m. (a) Image construction: The image nomenclature is the same as shown in Fig. 3. (b) $p(t)$ at $r=1.16$ m. The first arrivals are d, 1, and $1'$. The travel times for 2 and $2'$ are the same because $\theta=\theta_0$ and the pressures add. The next arrivals are 3 and $3'$. The arrivals 4 and $4'$ add because their travel times are the same. The diffraction arrival times for the first few diffractions are indicated by the symbol Δ. The amplitudes of diffractions SA, $1'A$, etc. are less than 1/30 of the amplitude of the direct arrival and do not show on this figure.

curvatures of the wall board construction of the wedge planes change the raypaths.

IV. SOURCE AND RECEIVER ON SHELF: SIMULATION 2

Figure 7 shows the experimental arrangement and image construction. The image construction is the same shown on Fig. 3. The pattern of arrivals is shown by the first few arrivals. The pairs of arrivals 2 and $2'$, 4 and $4'$, etc. add because their travel times are the same ($\theta=\theta_0$). This image construction does not include the diffractions from A and A$'$. However, the arrival time of the raypath S–A–R (or SA for short) is indicated. The theoretical results, as shown in Fig. 8, are from the convolution of the impulse response and the source function (Fig. 5). Experimental data are shown in Fig. 9. The source impulse is at $t=0$. Wedge angle $\theta_W=10.8°$. Source and receiver angles are $\theta=\theta_0=6°$. The source range from the apex O is OS=0.53 m. The receiver ranges are OR=1.16 m to 1.24 m. (a) Image construction: The image nomenclature is the same as shown in Fig. 3. (b) $p(t)$ at $r=1.16$ m. The first arrivals are d, 1, and $1'$. The travel times for 2 and $2'$ are the same because $\theta=\theta_0$ and the pressures add. The next arrivals are 3 and $3'$. The arrivals 4 and $4'$

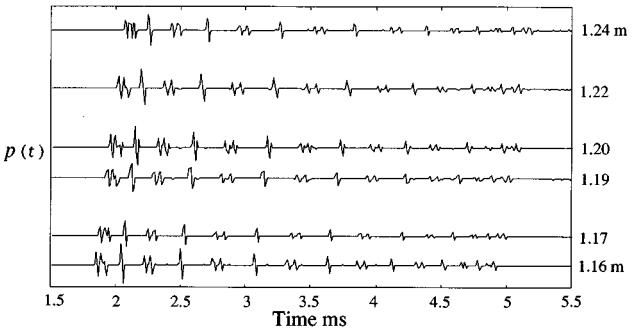


FIG. 8. Theoretical simulation 2. The geometry and an impulse response are shown in Fig. 7. Times are measured from the initiation of the spark at $t=0$. The impulse responses were convolved with the spark signal (Fig. 5). Theoretical simulation 2: The geometry and an impulse response are shown in Fig. 7. Times are measured from the initiation of the spark at $t=0$. The impulse responses were convolved with the spark signal (Fig. 5). Identification of arrivals are in Fig. 9.

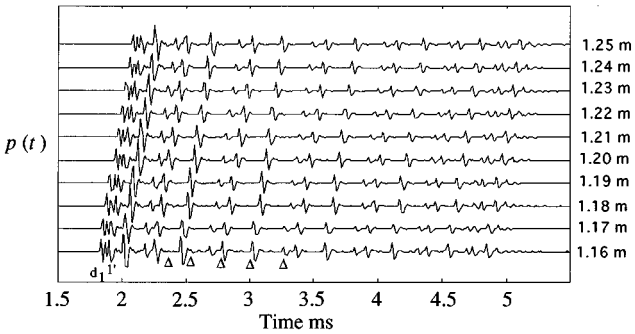


FIG. 9. Experimental data for simulation 2. The experimental parameters are given in Figs. 7 and 8. Time is measured from the initiation of the spark at $t=0$. Some of the arrivals are identified above the top trace. The arrival times of diffractions are indicated by the symbol Δ.

add because their travel times are the same. The diffraction arrival times for the first few diffractions are indicated by the symbol Δ. The amplitudes of diffractions SA, 1′A, etc. are less than 1/30 of the amplitude of the direct arrival and are not shown in this figure.

The theoretical transmissions to a set of receiver positions are shown in Fig. 8. The set of impulse responses were convolved with the spark signal (Fig. 5). Individual arrivals are from the ''odd'' image numbers 1 and 1′, 3 and 3′, etc. The ''even'' numbered image arrivals 2 and 2′, 4 and 4′, etc. have the same travel times and add to give almost double amplitudes. The geometry and an impulse response are shown in Fig. 7. Times are measured from the initiation of the spark at $t=0$. The impulse responses were convolved with the spark signal (Fig. 5). Identification of arrivals are in Fig. 9.

A set of experimental pressure transmissions are shown in Fig. 9. In comparison with Fig. 8, the arrival times and ''character'' of the traces match the theoretical transmissions. However, the sound pressure for the 3 arrival is less than that of the 3′ arrival. The same pattern continues for the 5 and 5′ arrivals, etc. This is believed to be caused by a small deformation or curvature of the dry construction (less than 1 cm in 120 cm). As expected, the sum pressure amplitudes of the 2 and 2′ arrivals are larger than 3, etc. Because of the geometry, the diffractions from *A* have nearly the same travel times to all receivers and would be identifiable if their amplitudes were larger than the background noise. Nu-

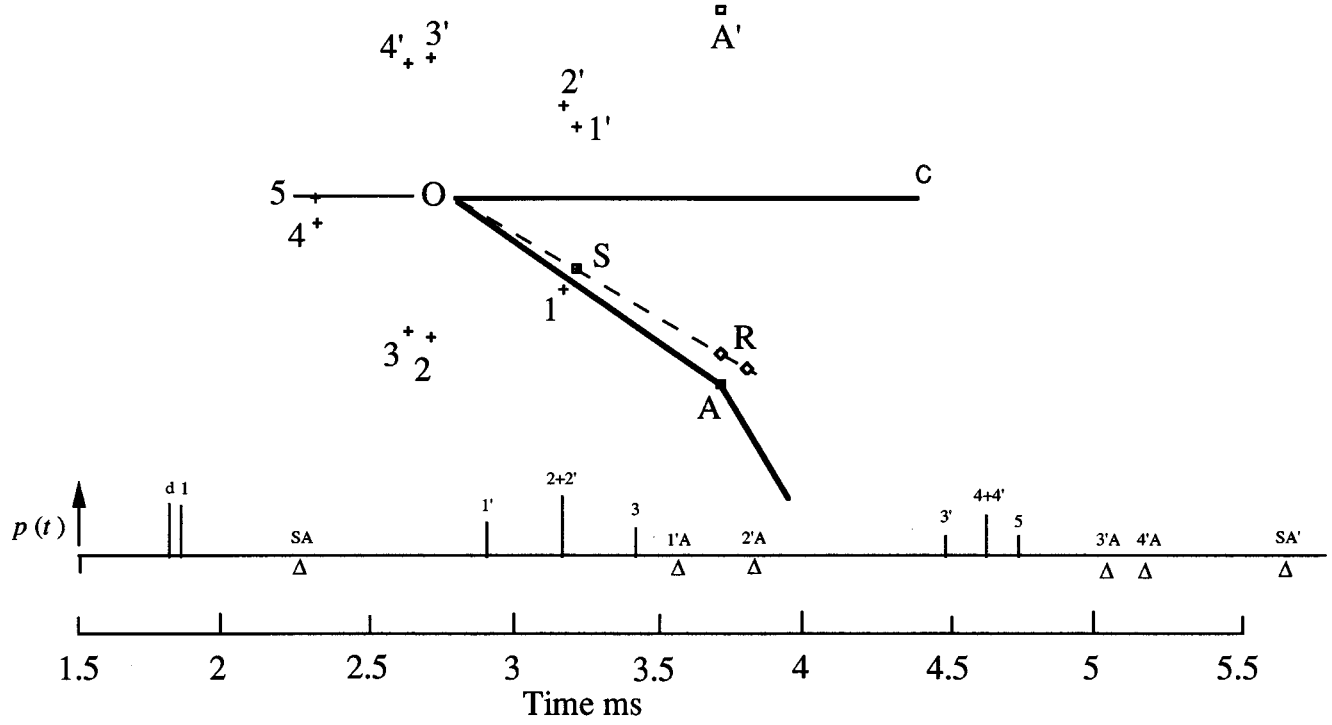


FIG. 10. Image construction and $p(t)$ for simulation 3. Wedge AOD is 35° and O–A=1.22 m. Wedge BDC is 59°. Acoustic parameters are $\theta_w=35°$, $\theta=\theta_0=30°$, O–S=r_0=0.53 m, and O–R=r=1.165–1.255 m. The impulse transmission $p(t)$ is for $r=1.165$.

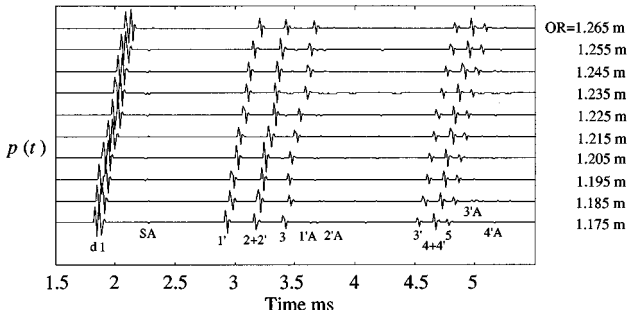


FIG. 11. Theoretical transmission for simulation 3. Time is measured after the initiation of the spark. Identifications are shown in Figs. 10 and 12.

merical computations give peak amplitudes of the diffractions as being less than 1/30 of the direct arrival. The diffraction arrivals are below the noise level. The experimental parameters are given in Figs. 7 and 8. Time is measured from the initiation of the spark at $t=0$. Some of the arrivals are identified above the top trace. The arrival times of diffractions are indicated by the symbol Δ.

V. SIMULATION 3

The model simulation for a steep shelf is shown in Fig. 10. This profile simulates the transmission down a stepper shelf. The receivers are near the shelf break at A. Compared to the shallow slope shown in Figs. 7–9, the steep slope has reduced the number of image arrivals. The times of diffraction arrivals are indicated by the Δ's. Wedge AOD is 35° and OA=1.22 m. Wedge BDC is 59°. Acoustic parameters are $\theta_W=35°$, $\theta=\theta_0=30°$, OS$=r_0=0.53$ m, and OR$=r$ $=1.165$ to 1.255 m. The impulse transmission is for r $=1.165$.

The convolution of the impulse response and the spark signal are shown in Fig. 11. The identification of arrivals follows our usual notation. Again the same source and receiver angles [$\theta=\theta_0=30°$] cause some of the arrivals to have the same travel times, i.e., 2, 2′ and 4, 4′. Time is measured after the initiation of the spark. Identifications are shown in Figs. 10 and 12. The reflection coefficients $\mathscr{R}_{10}$ and $\mathscr{R}_{12}$ are 0.9.

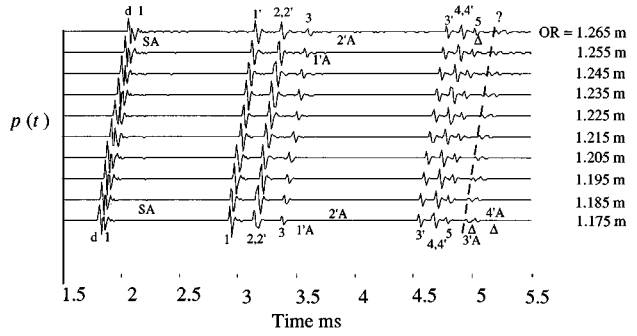


FIG. 12. Experimental transmissions for simulation 3. The identification follows the nomenclature used previously. The diffraction arrival SA is indicated. The other very small arrivals are not identified. The arrival labeled ''?'' is probably an artifact from the coaxial spark.

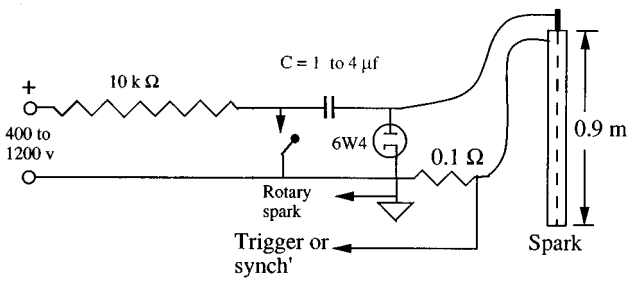


FIG. 13. Coaxial spark.

The transmission experiment for the steep shelf, simulation 3, is shown in Fig. 12. It is easy to identify and follow all of the image arrivals. It is difficult to identify all of the diffraction arrivals. The diffraction arrival SA is evident as a small negative bump for the transmissions to the receivers at 1.185–1.235 m. The others are too small. The dashed line labeled ''?'' is believed to be an artifice from the coaxial spark and is a delayed ''transmission' from the source. The slope of the dashed line is the same as d for the direct arrivals from the spark. The identification follows the nomenclature used previously. The diffraction arrival SA is indicated. The other very small arrivals are not identified.

VI. CONCLUSIONS

The conclusions are quite simple. The raytrace or image arrivals account for practically all of the acoustic energy. As given in the Biot–Tolstoy theory, one gets a finite number of image arrivals. The number of arrivals depends on the wedge angle, source angle, and receiver angle. Diffraction arrivals were small and usually below the background noise. The image-raypath constructions gave impulse responses and the convolutions of these impulse responses gave a set of theoretical transmissions. Within small amplitude differences, the theoretical and experimental transmissions matched.

ACKNOWLEDGMENTS

Research support came from the following: the Center for Polar Research and the Department of Geology and Geophysics of the University of Wisconsin; the Office of Naval Research, Roger Meredith (N00014-91-5-6004) and Mohsen Badiey (N00014-89-J-1515); and the Ocean University of Qing Dao.

APPENDIX: COAXIAL SPARK

The coaxial spark was made for research on seismic models.[16] The spark ran at about 1 spark/s for hours and transmitted very nearly the same waveforms and pressure amplitudes. The peak sound pressures are large and ultra quiet acoustical laboratories were not needed.

The construction of the spark is shown in Fig. 13. Insulation tubing, ''spaghetti,'' was heated and stretched to reduce the diameter of an end. With a thread, the insulation was pulled through the 3.18-mm o.d. (0.125 in.) stainless steel tube. The center conductor was straight #14 steel ''piano'' wire. After assembly, the spark end was ground square

and smoothed with a fine file. To prepare a new sparker, soft #1 graphite was rubbed over the end. After the spark has been running for a few minutes, the end chars and forms a resistance that helps to initiate the spark. We have also observed a wave that travels along the tube back to the upper end, reflects, and back down. Some of this wave appears as a lower amplitude radiation from the end of the spark. The length of the coaxial tube controls the duration of an experiment, about 5 ms for a 0.9-m tube. The spark also radiates ''radio'' waves. It helps to have the sparker system far from the computer. One can use an inductive pickup for the trigger.

We have used several combinations of supply voltage and capacitors. A rotary mechanical switch was used to initiate the discharge. The capacitor C is charged through R and the damping diode 6W4. The switch grounds one side of C and the other side of C is discharged through the spark. The mechanical switch can be replaced with a triggerable thyratron type of device.

[1] M. A. Biot and I. Tolstoy, ''Formulation of wave propagation in infinite media by normal coordinates with an application to diffraction,'' J. Acoust. Soc. Am. **29**, 381–391 (1957); I. Tolstoy and C. S. Clay, *Ocean Acoustics* (American Institute of Physics, New York, 1987), 2nd ed., Appendix 5.

[2] I. Tolstoy, *Wave Propagation* (McGraw-Hill, New York, 1973), Chap. 8.

[3] C. S. Clay, *Elementary Exploration Seismology* (Prentice-Hall, Englewood Cliffs, NJ, 1990), Sec. 14.1.2.

[4] C. Feuillade and C. S. Clay, ''Broadband source imaging in a shallow water wedge by an array of receivers,'' J. Acoust. Soc. Am. **96**, 501–514 (1994).

[5] H. Medwin and C. S. Clay, *Fundamentals of Acoustical Oceanography* (Academic, Chestnut Hill, 1997), Sec. 11.8.

[6] G. M. Jebsen and H. Medwin, ''On the failure of the Kirchhoff assumption in backscatter,'' J. Acoust. Soc. Am. **72**, 1602–1611 (1982).

[7] H. Medwin, E. Childs, and G. M. Jebsen, ''Impulse studies of double diffraction: A discrete Huygens interpretation,'' J. Acoust. Soc. Am. **72**, 1005–1013 (1982).

[8] S. Li and C. S. Clay, ''Sound transmission experiments for an impulsive source near rigid wedge,'' J. Acoust. Soc. Am. **84**, 2135–2143 (1988).

[9] C. S. Clay, D. Chu, and S. Li, ''Specular reflections of transient pressures from finite width plane facet,'' J. Acoust. Soc. Am. **94**, 2279–2286 (1993).

[10] S. Li, D. Chu, and C. S. Clay, ''Time domain reflections and diffractions from facet wedge constructions: Acoustic experiments including double diffractions,'' J. Acoust. Soc. Am. **96**, 3715–3720 (1994).

[11] J. P. Chambers and Y. H. Berthelot, ''Time-domain experiments on the diffraction of sound by a step discontinuity,'' J. Acoust. Soc. Am. **96**, 1887–1892 (1994).

[12] J. R. Preston and W. A. Kinney, ''Monostatic and bistatic reverberation results using linear frequency modulated pulses,'' J. Acoust. Soc. Am. **93**, 2549–2565 (1993).

[13] C. S. Clay and W. A. Kinney, ''Numerical computations of time domain diffractions from wedge and reflections from facets,'' J. Acoust. Soc. Am. **83**, 2126–2133 (1988).

[14] H. Medwin, ''Shadowing by finite noise barriers,'' J. Acoust. Soc. Am. **69**, 1060–1064 (1981).

[15] M. J. Buckingham, ''Theory of three dimensional acoustic propagation in a wedge like ocean with a penetrable bottom,'' J. Acoust. Soc. Am. **82**, 198–210 (1987).

[16] C. S. Clay and H. McNeil (Morris), ''An amplitude study on a seismic model,'' Geophysics **20**, 766–773 (1955).

Sound generation and air entrainment by breaking waves in the surf zone

Grant B. Deane
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0238

(Received 5 November 1996; revised 14 July 1997; accepted 22 July 1997)

This paper presents the results of acoustic and optical measurements of individual breaking waves in the surf zone. Two hydrophones, horizontally separated and deployed in 2 m of water off La Jolla Shores beach, California, were used to measure the source spectrum of breaking surf, and characterize propagation through the surf zone over length scales of order 10 m. The acoustic data show an acoustically active region within a wave which propagates shoreward. The production of sound within the active region is associated with the formation of bubble plumes. Above 500 Hz, the sound is consistent with the radiation from individual bubble oscillations. Pictures were taken of the bubble plumes formed beneath the breaking surf, providing estimates of the plume size, and bubble size distribution and void fraction within a plume. The density of bubbles scales as $a^{-2.5}$ for $a<1$ mm and $a^{-4.5}$ for $a>1$ mm, where a is bubble radius, and total void fractions of 0.3–0.4 were measured. Theoretical calculations show that radiation from bubbles within a plume is strongly damped, and only a thin shell of bubbles on the exterior of the plume contributes to the wave noise. A ''moving bubble sheet'' source model for the wave noise is presented which shows fair agreement with the acoustic measurements.
[S0001-4966(97)02611-8]

PACS numbers: 43.30.Nb [SAC-B]

INTRODUCTION

The study of ambient noise in the open ocean and coastal waters is a well-established field. Hundreds of papers have been published on the subject during the years since World War II (the reader is referred to Urick[1] for a survey). The acoustics of breaking waves in the open ocean, a closely related topic, is also a maturing field summarized in a recent review article by Melville.[2] Little of this extensive work has addressed the properties and sources of ambient noise in the oceans' margins, within a few kilometers of the shore. Although a number of the acoustic studies of laboratory breaking waves (see, for example, Melville and colleagues,[3–7] Medwin and Daniel,[8] and Papanicolaou and Raichlen[9]) are directly relevant to the surf zone, few field measurements of the noise field in or close to surf have been published.

Breaking waves are the dominant source of wind-driven noise in the open ocean, a fact recognized by Knudsen *et al.*[10] five decades ago. The generating mechanism remained a source of speculation until relatively recently; it is now generally accepted that the noise radiated by deep water waves breaking under the influence of wind, above a few hundred Hz, is sound of ringing bubbles created when the wave breaks (an understanding largely reached during the course of two relatively recent conferences[11,12]). The underwater sound generated by waves breaking on the sea shore is known to contribute to the ambient noise in coastal waters. Surf noise was measured up to 9 km from the shore in Monterey Bay, California during an experiment conducted by Wilson *et al.* In one of the few published measurements of surf noise, Wilson *et al.*[13] determined a seaward versus shoreward anisotropy (energy from the shoreward direction predominating) in the ambient noise field in the frequency range from 20 Hz to 700 Hz. The anisotropy, which depended on range from the beach, acoustic frequency and wave height, was 10 dB at 300 Hz and 9 km from the shore during very heavy surf. They concluded that a component of the ambient noise field in the near shore region is due to surf. Earlier studies by Bardyshev *et al.*,[14] and Zakharov and Kirshov[15] support this conclusion.

The region of the surf zone between the range at which waves first break and the shore line can be divided into an outer, or ''transition,'' region and an inner region as described by Svendsen.[16] In the transition region, which covers a horizontal range of approximately 8–10 water depths after the initial wave break,[16] there is a rapid evolution in the wave shape. The inner region is shoreward of the transition region; waves propagating in this region evolve more slowly and have many features in common with bores.[17] Wiegel[18] identifies three categories of breaking waves in the surf zone, each of which exhibits a characteristic breaking. The categories are, in increasing frequency of observation on beaches:[19] surging, plunging, and spilling. The waves studied during the experiments described here were observed in the transition region of the surf zone. They were, in category, in the transition between spilling and plunging breakers.

A series of eight images illustrating a breaking event in the transition region is shown in Fig. 1. The first image was taken as the wave began to break with subsequent images taken at 1-s intervals. The air entrained in the leading region of the wave changes the optical properties of the water, and from the surface appears as a white, foamy area. The first four images show the wave evolution in the transition region; by roughly the fifth or sixth image, the wave has reached the inner region and propagates much like a bore.

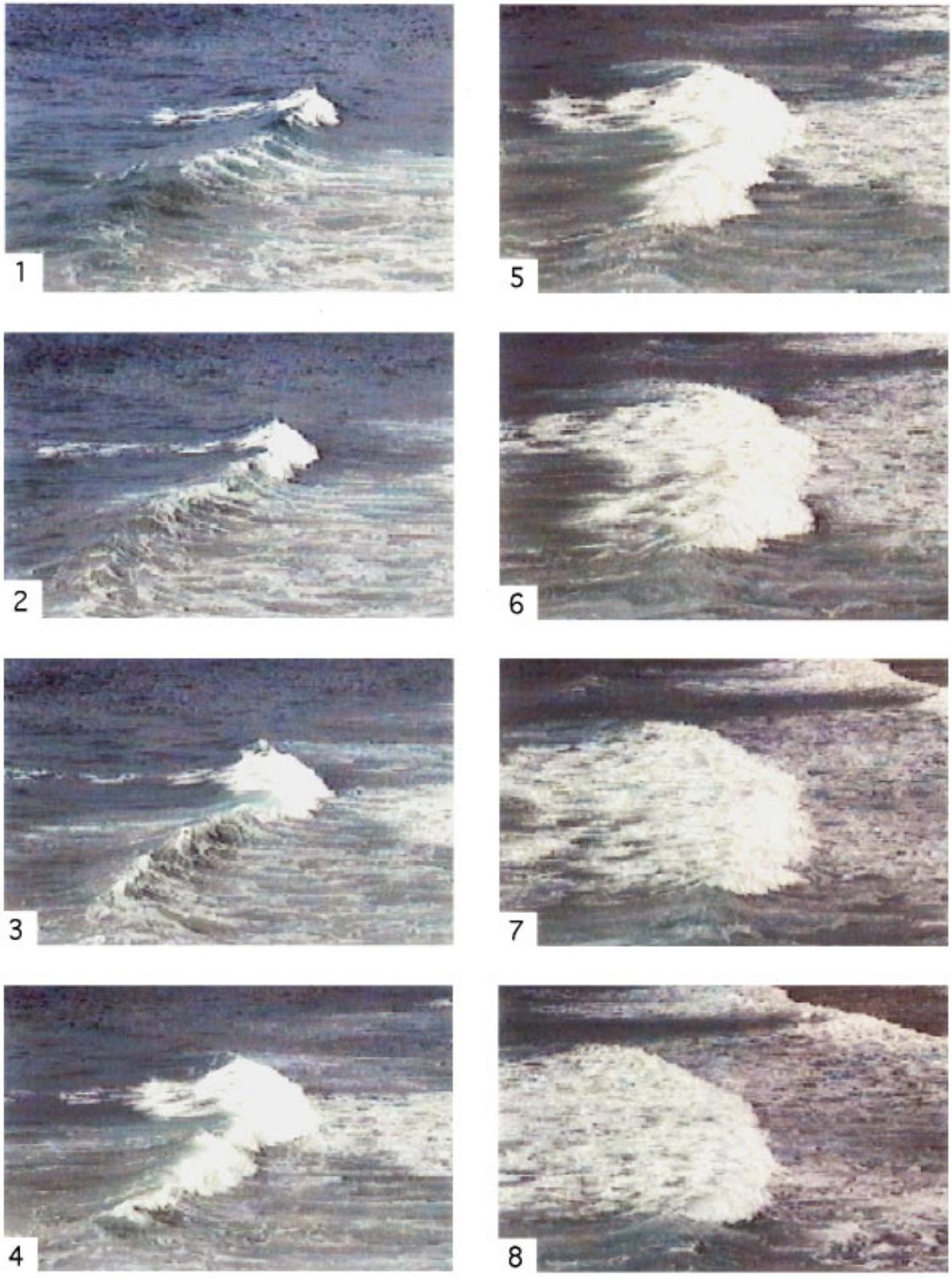


FIG. 1. A sequence of eight images of a breaking wave in the surf zone. Successive images were taken at 1-s intervals.

Given the connection between the near-shore ambient noise field and breaking surf, a central question is to determine the character of the surf noise and its seaward propagation from the surf zone. Surf source level measurements have been made by Stewart and Wilson,[20] who estimated the average power spectral density due to heavy surf at Fort Ord beach, Monterey. The estimates were formed by accounting for the attenuation due to propagation effects on surf noise measurements made 8.5 km from the beach. The noise spectrum observed some distance from the surf is a superposition of all wave breaking events audible at the measurement point, and includes the combined effects of the source spectrum and transmission loss through the surf zone. The sound of many breaking waves contributed to Stewart and Wilsons' measurements, and their source levels represent a spatial and temporal average of the noise radiated per meter of shore

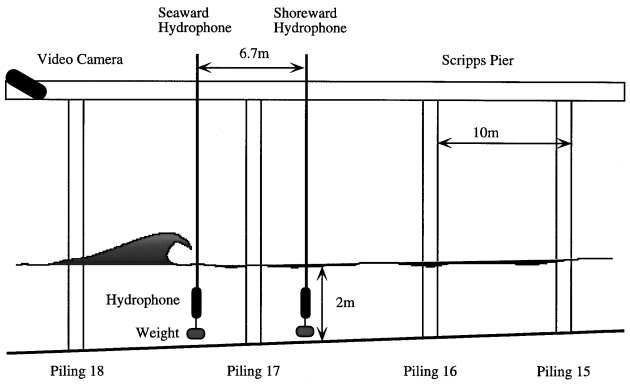


FIG. 2. The hydrophone and video camera configuration for the acoustic experiment. Breaking waves move from left to right.

line. In this paper we are concerned with the sound production of individual breaking waves in the surf zone. We have taken a different approach from Stewart and Wilson, and present direct measurements of the acoustics and air entrainment of single breakers.

Section I describes an experiment to measure the underwater sound of individual breaking waves by means of hydrophones placed within a few meters of the surf. Section II describes a complementary optical experiment to obtain qualitative data on the gross features of air entrainment, and a quantative measure of the bubble size distribution within a bubble plume. Sections III and IV present a review of possible underwater sound source mechanisms associated with surf, including free-bubble oscillations, collective bubble oscillations, jet splashes, sediment disturbance, hydrodynamic pounding of the sea floor, turbulence, and infrasonic sources of ''pseudo-sound.'' In Sec. V a model is developed for the sound which accompanies the creation of a bubble plume. This model is subsequently used in Sec. IV, along with a canonical environmental geometry, to compute the sound field in the surf zone wave guide and establish the first-order physics of propagation within a few meters of the waves. Once developed, the model can be used to back-out the effects of the surf zone waveguide on the sound field and establish an estimate for the breaking wave source spectrum. This is done in Sec. VII. Section VIII presents the concluding remarks.

I. EXPERIMENT I: ACOUSTIC MEASUREMENTS

During September 1995, a series of underwater sound and surface video recordings were made of individual breaking waves in the surf zone. The experiments were conducted from the pier at Scripps Institution of Oceanography, which extends 400 m into the littoral zone of La Jolla Shores beach, San Diego. The sensor array consisted of two broadband hydrophones separated horizontally, and a video camera pointed at the ocean surface (see Fig. 2). The hydrophones were suspended from the side of the pier with cables and placed 6.7 m apart and 1 m below the mean surface level of the water. Simultaneous measurement of the sound field at two horizontally separated locations allowed propagation of wave noise to be studied over short length scales. Despite weights attached to each cable, the hydrophones tended to be displaced along the direction of wave motion as waves broke. Analysis of the video data while the camera was pointed directly above the seaward hydrophone allowed an estimate of the hydrophone's shoreward velocity, which was 2.5 m/s during the passage of the larger waves. The hydrophone motion occasionally resulted in jerks on the hydrophone cable which caused signal contamination, particularly during the passage of larger waves. Events affected in this way were removed from the data set. Optical measurements of the bubble plumes formed beneath breaking waves were made after the acoustic experiment, and are described in detail in Sec. II. Comparable mean wave heights, wind speeds, and directions on the days of the two experiments yielded breaking waves of a very similar type. Photographs of the bubble plumes created beneath the waves [see Fig. 4(a) and (b), for example] show that the plumes extended between 0.5 m and 1 m below the surface, and consequently the hydrophones were beneath the plumes for all but the largest events.

The video data were recorded on a CCD camera using the Hi-8 video format. The camera was deployed in two configurations, pointed vertically downward above the seaward hydrophone and at an angle so that both hydrophones were visible. The acoustic data were recorded with a digital audio tape, providing a dynamic range of 96 dB and bandwidth of 20 kHz. The video and acoustic data were synchronized by recording a few minutes of the tape time display before deploying the camera over the side of the pier. This allowed the two instruments to be synchronized to within one video frame, or 33 ms.

The experiment lasted for 3 h, during which time in excess of 100 breaking events were observed. The measurement site was chosen to coincide with the seaward edge of the transition region. The wave fronts were parallel to the local depth contours, and tended to break simultaneously along the front. The largest waves broke a few meters before the seaward hydrophone with smaller waves breaking in shallower water, closer to the shore. The pier pilings provided reference points in the video recording for the offshore range. The pilings in view of the camera are respectively numbered 15–18 running from east to west (see Fig. 2). A designation of class 1 was given to waves which overturned before piling 18, class 2 for waves which overturned between pilings 18 and 17, and class 3 for waves which overturned at piling 17. Waves that broke shoreward of piling 17 were not clearly distinguishable from the background surf and pier noise. Visual observation of the waves showed that they were in the transition region between spilling and plunging breakers.

The U.S. Army Core of Engineers maintains a wave height meter and a meteorological station on Scripps pier which recorded the significant wave height and period, respectively, as 0.9 m and 7 s throughout the experiment. The wind speed varied between 3 and 5 m/s, blowing from the west (247–258 deg). The water temperature was 18 °C and the beach slope approximately 0.025 rad (1.4 deg). The water depth at the observation area was measured by dropping a line over the side of the pier, and was roughly 2 m.

The sea bed at La Jolla Shores beach consists of a fine,

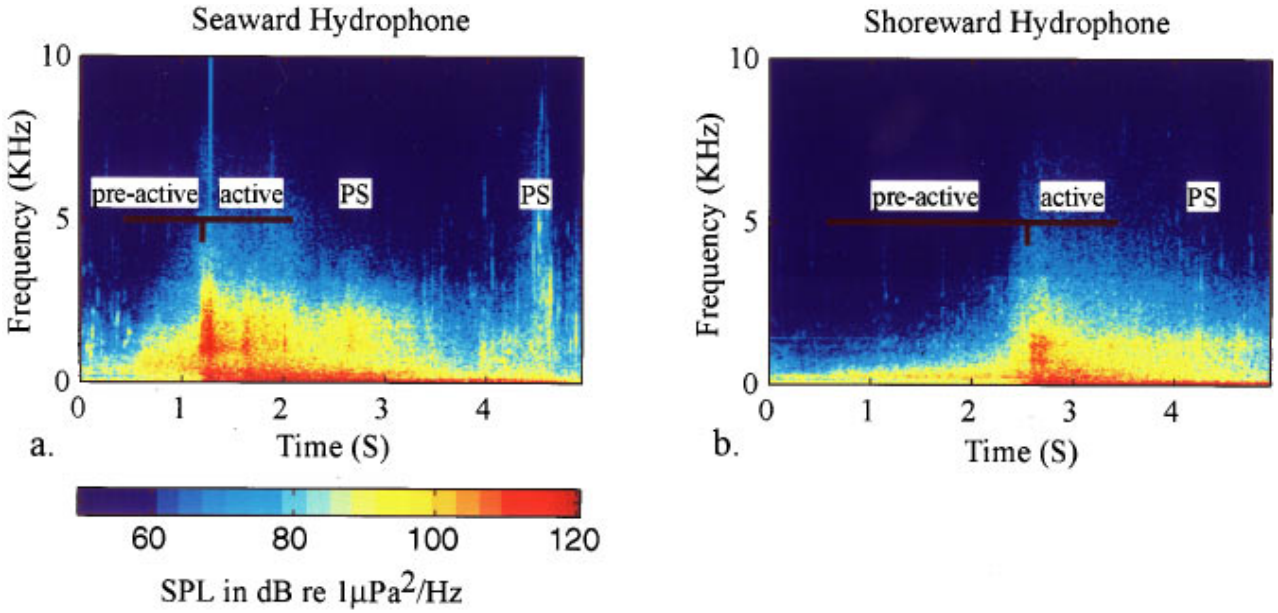


FIG. 3. Frequency versus time plots of the noise produced by an energetic breaking event recorded by (a) the seaward hydrophone and (b) the shoreward hydrophone. The periods labeled ''pre-active,'' ''active,'' and ''PS'' are discussed in the text.

relatively uniformly grained lithogenous sand. The grain sizes have been measured, and range in diameter from 0.2 to 0.5 mm.

A. A wave event spectrogram

Spectrograms of the sound produced at the seaward and shoreward hydrophones by a class 1 event are shown in Fig. 3(a) and (b). The designation ''PS'' means piling splash, and indicates an acoustic event caused by the breaking wave interacting with a pier piling. These interactions are discussed in more detail in the Appendix.

The video recording of the wave (not presented here) shows a clear sequence of events which correspond to the pre-active and active periods marked on the spectrograms. The pre-active period began when the wave first overturned and lasted until the breaking region reached the hydrophone. The active period began when the leading edge of the breaking region reached the hydrophone and lasted until the breaking region passed by. By the term ''breaking region'' we mean the region which extends from the leading edge of the breaking wave to the wave crest. The mixture of air and water in this region gives it a white, foamy appearance (see Fig. 1). Each of the events described above is visible in the video recording with the exception of the start of the wave break, which occurred outside the field of view.

The active period begins with a sudden increase in acoustic energy coincident with the passage of the breaking wave over the hydrophone. The increase in intensity with the arrival of the breaking region is consistent with the proximity of the hydrophone to the sources of the sound observed during the pre-active period. Clearly, and not surprisingly, the breaking region of the wave is acoustically active. Since the term ''breaking'' refers to a hydrodynamic process, we introduce the term ''acoustically active region,'' or just ''active region'' to describe the volume of the breaking wave which is producing sound.

There is a time lag of 1.4 s between the beginning of the active period of the seaward and shoreward hydrophones. This lag is the travel time of the active region over the 6.7 m between the two sensors, and corresponds to a wave speed of 4.8 m/s. Simple shallow water wave theory predicts a speed of $u=\sqrt{gh}$, where $g=9.8$ m/s is the acceleration due to gravity and h is the water depth. Taking $h=2$ m yields a theoretical speed of 4.4 m/s, which is in reasonable agreement with the results inferred from the spectrograms.

The pre-active period begins when the sound of the approaching wave rises above the background surf noise and ends when the active region reaches the hydrophone. Because the active region may be 4 water depths or more away from the hydrophone at the beginning of the pre-active period, the acoustics during this time are a combination of the source spectrum of the active region and the transmission properties of the wave guide.

II. EXPERIMENT II: AIR ENTRAINMENT

The acoustic experiment gave a clear indication of sound sources in the breaking region, an observation which motivated an additional study to examine the air entrainment process occurring there. Breaking waves in the open ocean which form whitecaps create an ensemble of bubbles below the surface. The current knowledge of the dynamics and acoustics of this process have recently been summarized in a comprehensive review of the role of surface wave breaking in air–sea interaction by Melville.[2] The temporal evolution of a bubble ensemble after its creation is a complicated problem.[21,22] We will not consider this process in detail, but adopt the terminology used by Carey *et al.*[23] based on Monahan and Lus[24] model for the evolution of a bubble ensemble. Carey *et al.* use two terms depending on the age of the ensemble, ''plume'' for the first few seconds, and ''cloud'' thereafter.

Like ocean whitecaps, breaking waves in the surf zone form bubbles in the water column. The goal of the air entrainment experiment was to characterize the size and spatial distribution of bubble plumes and clouds after the passage of a breaking wave. To this end, pictures were taken of the plumes formed beneath the wave and the foam patterns formed behind the wave. The foam patterns were imaged with the same CCD video camera used to record the breaking waves during the acoustic experiment. The photographs of the plumes and clouds were taken with an underwater camera. Two sets of photographs were taken on separate days, the first to image the size and distribution of the plumes formed beneath the breaking region and the second to estimate the bubble size distribution within the plumes. The second set of images were taken with an extension tube attached to the camera to provide both magnification and a narrow depth of field.

During the 1970s, photographic methods were used by both Medwin[25] and Johnson and Cooke[26] to measure the bubble size distribution in coastal waters. Subsequent acoustic measurements by Breitz and Medwin[27] and laboratory studies by Baldy[28] showed that the optical method used by Johnson and Cooke was limited in accuracy for bubbles smaller than 60 μm radius. Any corresponding limitation in our study will not affect our results as we have restricted attention to bubbles greater than 100 μm.

Taking detailed photographs beneath breaking waves in the surf zone is challenging. A suitable event may occur at the camera location once every 2–5 min, providing a 1-s opportunity for an image. In addition, the creation of a bubble plume is, by necessity, accompanied by a large wave breaking directly overhead. Smaller waves are less energetic, but break nearer shore in shallower, turbid water and yield poor quality photographs. The slides presented here were taken by Dr. Dale Stokes, an experienced underwater photographer who captured some high quality images. However, because of the difficulties just outlined, our knowledge of the time and location of the images is imprecise; the photographs were taken in random locations within the first 0.5 m of the water column within a second or two of the formation of the plumes.

A. Images of bubble clouds

Two images of bubble ensembles formed beneath 1-m waves breaking off La Jolla Shores beach on 11 June 1996 are shown in Fig. 4(a) and (b). The wave was propagating from left to right in Fig. 4(a) and the view is looking upward in Fig. 4(b). The diver in the first figure gives an idea of the length scale. The air entrained as the wave breaks reduces the optical transmittance of the water, and regions of high air concentration appear as dark patches in the photographs. It is clear from the distribution of dark patches that the entrained air is packed into cells of bubbles. As the photographs were taken within a second or so of the wave passage, the bubble concentrations can be classified as ''plumes.''

The two bubble plumes directly above the diver in Fig. 4(a) are about 1 m long and 0.5 m high. Depending on how one defines the cell boundaries, there are eight or more plumes in the segment of wave shown in Fig. 4(b). Examination of a number of images like Fig. 4(a) and (b) indicate that the plumes are not randomly distributed in space, but have a propensity to be packed into lines parallel to the wave front. This conclusion is supported by observations of the foam patterns formed behind the wave.

Once created, a bubble plume begins to dissipate through a process of dissolution, diffusion, and degasing.[21] As bubbles degas and rise to the surface, they create foam which traces the spatial distribution of the bubble plumes and clouds below. Motion of the surface tends to destroy the foam patterns, which remain clear for a few seconds after the passage of the wave. Figure 5(a) and (b) shows the patterns generated by two breaking waves in the transition and inner regions. Lines of foam parallel to the wave fronts are clearly visible. These lines are a strong indication that the plumes and clouds generating the foam are also packed into lines.

The formation of bubble cells packed into lines is consistent with the existence of periodic flows associated with air-filled regions within the breaking wave. It is intriguing to note that such flows have been observed in laboratory studies. Lin and Hwung[29] have measured the turbulent flow of plunging breakers in a wave flume. By illuminating fluorescent-coated tracer particles, the authors determined the water velocity field in horizontal and vertical planes during a wave breaking event. They showed that flow in the transition region of a (plunging) breaking wave is cyclic, and consists of a series of vortex systems parallel to the leading edge of the breaking wave. As each vortex is formed, it entrains air which is eventually released as a cloud of bubbles behind the wave. The authors also report that the flow field is not uniform along a vortex, but forms a complicated, three-dimensional structure. Assuming that the laboratory measurements translate to the surf zone, the air entrainment mechanism observed by Lin and Hwung explains the intermittent nature of bubble cloud formation and the tendency of the bubbles to form lines parallel to the breaking wave.

B. The bubble size distribution

Images of the individual bubbles within plumes are shown in Fig. 6(a) and (b). Two sets of measurements were made on different days, providing a total of 40 usable slides. The dimensions of the field of view are 7 cm$\times$4.5 cm with a 0.9-cm depth of field, yielding a 28-cm^3 volume of water imaged per slide. The bubble ensembles were illuminated with an underwater strobe held above the camera as they were photographed. For reasons given earlier, we have only approximate information on the time that the pictures were taken, and the depth and orientation of the camera relative to the bubble ensemble.

A visual scan of the usable frames showed a considerable variability in the size and spatial distribution of the bubbles. A few images showed bubbles so densely packed that they had the appearance of a solid mass. The rest of the images varied in appearance from a relatively uniform spatial distribution with bubbles no larger than about 600 μm to swirls and clumps of millimeter-sized bubbles.

Two size distributions for the bubbles within an average plume have been obtained by analyzing four of the photo-

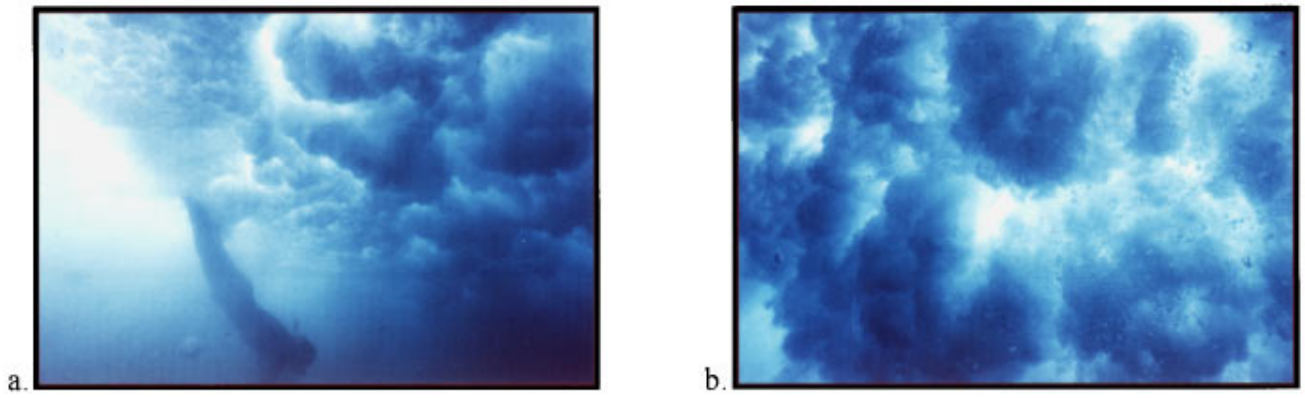


FIG. 4. Underwater photographs of the bubble plumes created by breaking surf in the transition region. The pictures were taken within roughly a second of the passage of the wave. Slides (a) and (b) show the plumes from two different breaking events.

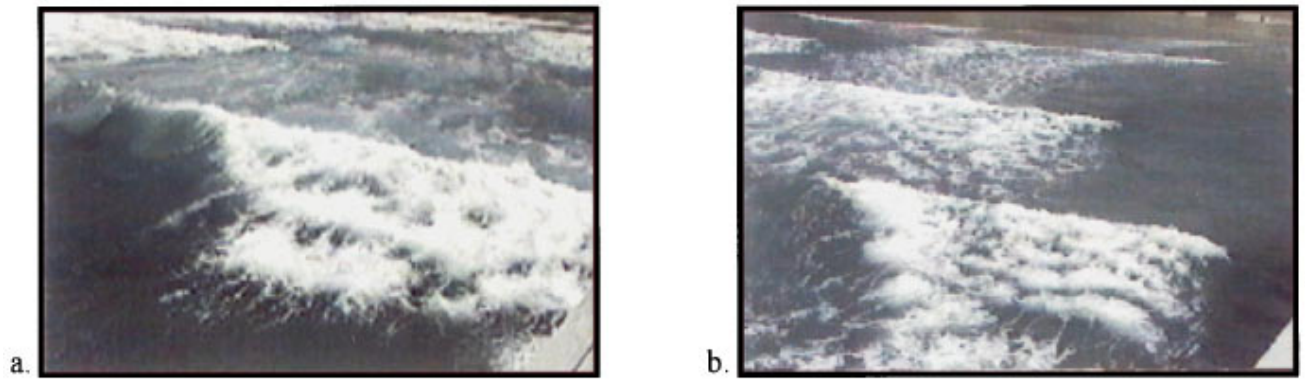


FIG. 5. Video frames of the foam lines created behind breaking waves in (a) the transition region and (b) the inner region. The foam patterns approximately trace the spatial distribution of the bubble plumes in the water column. Individual bubble plumes surrounded by bubble-free water can be seen.

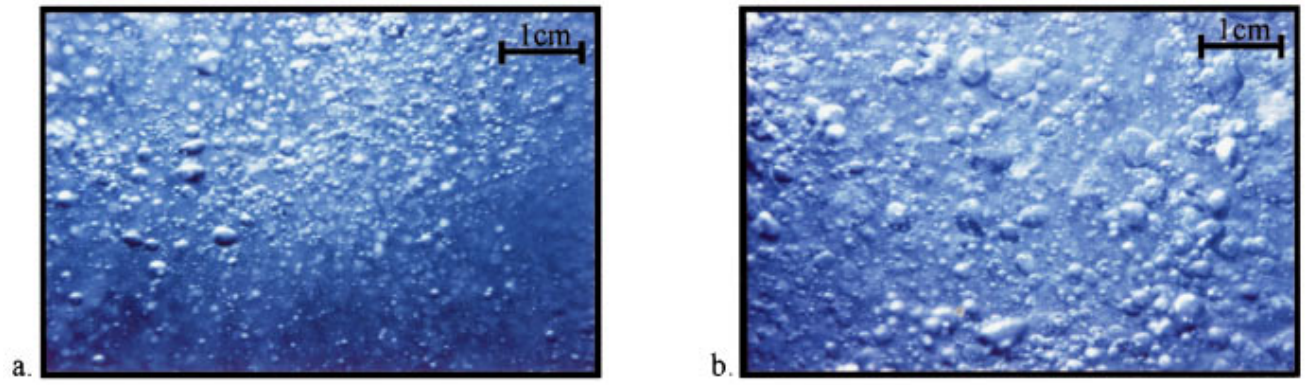


FIG. 6. Underwater photographs showing a magnified view of the individual bubbles within a bubble plume. These and similar images formed the basis of the bubble size distribution estimate.

graphs. Three of the photographs were selected because they had relatively uniform distributions and contained large bubbles. The fourth photograph was selected because it did not contain any large bubbles. An explanation for this selection is given below. The slides were scanned at a resolution equivalent to 30 pixels per mm of water column and converted into monochrome computer files. The files were analyzed using a Macintosh Power PC running NIH Image 1.60. The cross-sectional areas of 7681 bubbles were obtained by drawing an ellipse around each bubble (bubbles larger than 1 mm often showed an elliptical cross section) and computing its area. The radius of the bubble was then taken to be the radius of a circle with the same area as the ellipse.

The number density of bubbles $n(a)$ is the number of bubbles per unit volume which have radius in the range a to $a+da$, and is therefore a function of bubble radius a and the chosen radius increment da (usually taken to be 1 μm). The number density distribution was computed from the size estimates by counting bubbles which lay in a chosen range of radii or radius bin, and normalizing the result to counts per unit volume, per micron of radius increment. Ninety percent of the radii measured lay between 50 μm and 1 mm, the remaining ten percent being greater than 1 mm. Logarithmically spaced radii bins were used above 500 μm to accommodate the low bubble numbers at large radii.

One of the limitations of size distribution estimates presented here is a systematic bias introduced in the smaller bubble count. Small bubbles which are within the camera's depth of field but obscured by larger bubbles are inadvertently eliminated from the measurement. The three images containing large bubbles are affected in this way, and for this reason bubbles less than about 300 μm in radius were intentionally excluded from the measurement. We will call the size distribution estimated from the three images the ''large bubble distribution.'' There are no large bubbles in the fourth image, making it more suitable for a second estimate of the bubble size distribution, including bubbles down to 100 μm in radius. Because the bubble number density for the small bubbles is very high (there are approximately 10 000 bubbles in the image), only $\frac{1}{4}$ of the image surface was analyzed. We will refer to this distribution estimate as the ''small bubble distribution.'' The large and small bubble distributions together span bubble radii from 100 μm to 10 mm.

The volumes analyzed were 85 cm^3 taken from three plumes for the large bubble estimate, and 7 cm^3 taken from a single plume for the small bubble estimate. Although the total plume volume and number of plumes analyzed are small, the photographs studied were representative of the entire data set, and the resulting number density is indicative of the distribution within a plume.

The two size distribution estimates are shown in Fig. 7. The roll-off in number density at 300 μm in the large bubble estimate is an artifact of the optical measurement technique. Smaller bubbles were observed in the images, but they were ignored because of the small bubble bias discussed above. The number density in the small bubble estimate peaks around 100 μm. A similar peak in number density at 150 μm has been found in a study of the bubbles generated by gently spilling laboratory waves by Medwin and Daniel.[8] It is intriguing to note that both the large and small bubble distributions show two slight peaks at 400 and 500 μm. The presence of the peaks in both distribution estimates indicates they are a repeatable feature of the distribution estimates. Medwin and Daniel also observed two minor peaks in their number density distribution at 500 μm and 750 μm. The large bubble distribution shows a transition in scaling law, from $a^{-2.5}$ for radii less than 1 mm to $a^{-4.5}$ for radii greater than 1 mm, where a denotes radius. One possible interpretation for the rapid roll-off in number density at large radii is that we are seeing the outer length scale of the bubble size. Two other possibilities for the relative scarcity of larger bubbles are that they rose to the surface too rapidly to be photographed, or exist only near the surface, outside the camera's field of view. A linear extrapolation of Thorpe's[21] summary of the ascent speed of ''clean'' bubbles indicates a 6-mm radius bubble rises at roughly 1.4 m/s. Thus a 6-mm bubble injected 0.5 m into the water column reaches the surface in 0.35 s, assuming its ascent is not affected by neighboring bubbles.

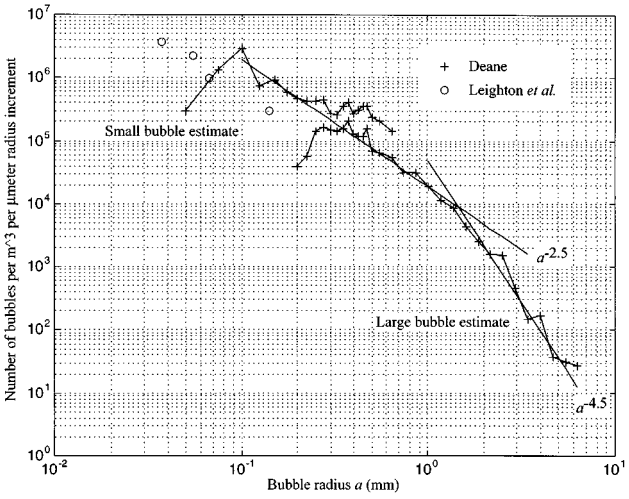


FIG. 7. The result of the bubble size distribution estimate based on the bubble plume photographs. Leighton *et al.*'s (Ref. 30) acoustic-based measurements in the surf zone are shown for comparison.

Measurements of the bubble size distribution in the surf zone recently made by Leighton *et al.*[30] are plotted in Fig. 7 for comparison with the optical measurements. Leighton *et al.* used a combination of acoustic techniques to determine number densities at four points between 40 and 110 μm. Their acoustic system was deployed at 1.5 m in 3-m-deep water in the presence of surging breakers. Our estimates are roughly a factor of 3 higher than those measured by Leighton *et al.* and agree favorably in level and slope to 100 μm. The roll-off in bubble density observed for radii less than 100 μm in the optical data may be due to the limited size resolution of the technique, although Medwin and Daniels'[8] acoustic laboratory data also show a roll-off in the number density of small bubbles. Further measurements are required to clarify the issue.

C. The void fraction

With the estimate of the bubble size distribution within a plume, we can calculate the average plume void fraction of

air. The void fraction of air within a region is defined as the ratio of the volume of air to the volume of the region, and is given by

$$\alpha \equiv \frac{V_p}{V} = \frac{4\pi}{3V} \int_{a_0}^{a_1} n(a) a^3 \, da, \tag{1}$$

where α is the void fraction, V_p is the volume of air within the chosen volume V, and a_0 and a_1, respectively, are the radius of the largest and smallest bubbles present within the volume. Since bubble volume scales as radius cubed, the scaling law for the size distribution determines whether large or small bubbles play a dominant role in determining the void fraction. Suppose the size distribution scales as $a^{-\gamma}$, where γ is a positive constant. If $\gamma > 3$ then small bubbles are important in determining the void fraction, whereas if $\gamma < 3$ it is the large bubbles which contribute most to the void fraction. For the distribution presented here, the bubbles around the transition in scaling law dominate the estimate, a fact that can be verified by computing the increment to the void fraction from bubbles which lie within a specified range of radii. The point is that the optical technique provides a representative estimate for the void fraction; the $a^{-2.5}$ scaling law for small bubbles relieves us from having to know the number of bubbles of small radius.

The void fraction is found by summing the bubble volumes and dividing by the volume of the bubble cloud imaged. Performing this calculation yields a void fraction of 0.34, one-third of which comes from bubbles with radii between 1 and 2 mm. Detailed measurements of the temporal and spatial distribution of the void fraction in bubble plumes generated by fresh water breaking waves in the laboratory have been made by Lamarre and Melville.[4] They used specially designed probes which measured the void fraction directly.[3] Void fractions as high as 0.54 were observed beneath 0.7-m waves immediately after breaking, a result which agrees favorably with the measurements presented here.

III. SOUND SOURCE MECHANISMS

The acoustic experiment presented in Sec. I A showed that breaking waves in the surf zone are noisy. Furthermore, the sources of sound, radiating from a few tens of Hz to 5 kHz or more, are located in the breaking region. The question addressed here is what the sources might be. Possible mechanisms include the sound associated with the entrainment of air, including free bubble oscillations and collective oscillations, splashes, sediment disturbance, ''surfseisms,'' turbulence, and infrasonic sources of ''pseudo-sound,'' which are nonacoustic (i.e., nonradiating) pressure fluctuations associated with the flow of water around a hydrophone.[31]

In a series of three abstracts published during 1960 and 1961, Saenger[32–34] presented experimental and theoretical considerations of the elastic waves generated in the sea floor by pounding surf. In an experiment conducted on the beach at Fire Island, New York, a correlation was found between surf activity and the vertical velocity of the surface sand layer near the surf. Saenger called the phenomenon ''surfseisms,'' presumably because of its similarity to the sea bed motion induced by nonlinear, wave–wave interactions known as microseisms. Although he ultimately concluded that the surfseism contribution to the ambient noise spectrum a kilometer offshore due to submeter sized waves is negligible, Saenger's work does not preclude a contribution to the sound spectrum directly beneath a breaking wave below 100 Hz.

Visual observations of the water surface and the sea floor behind 1-m breaking waves in La Jolla Shores beach showed sediment suspended in the water column, indicating that the sea bed had been disturbed by the passage of the wave. Thorne[35] has shown that sediment disturbance results in the generation of sound caused by colliding particles. The impulsive acceleration of sediment particles results in rigid body radiation, the spectral shape of which has the approximate form of a low-pass filter with a cutoff frequency depending on the particle size.[36] As the sediment grain size increases, the cutoff frequency decreases, and the power spectrum level increases.[35] Thus gravel disturbed by breaking waves makes a more significant contribution to the noise level at lower frequencies than sand, other factors being equal. Bardyshev *et al.*[14] have reported noise measurements around two pebbly ocean beaches and a rocky inland seacoast, and concluded that the primary acoustic source was the impact noise of pebbles agitated by waves. Measurements of the power spectrum levels[35] of the sound radiated by 0.5 kg of 160-μm particles (similar in size to those forming the sea bed of La Jolla Shores beach) rotated in a drum indicate that sediment disturbance, when it occurs, probably makes a significant contribution to the wave noise spectra presented here in the region of roughly 50–400 kHz. Measurement of the mass and velocity of sediment ejected by the wave are required before quantitative estimates of sea bed noise spectrum levels can be extrapolated from Thorne's work. We can, however, be reasonably confident that sediment disturbance is not a significant source of sound in the results presented here below 50 kHz.

Even the casual observer will discern that the wave breaking process in the transition region of the surf zone is energetic and accompanied by splashing. In the case of plunging breakers, the defining characteristic is a well-developed jet and subsequent splash as the wave first over turns. Splashes as sources of sound have been studied by Franz,[37] and more recently by Nystuen,[38] Pumphrey and Crum,[39] and Oguz and Prosperetti,[40,41] largely during the investigation of the underwater sound of rain. The splash of a drop on a fluid surface produces sound through two processes, the initial drop impact and radiation from any air bubbles which may be entrained. The sound of the initial impact increases with both drop size and velocity, but in total energy is usually much less than the sound energy radiated by an entrained air bubble.[37] Pumphrey and Crum[39] have studied the acoustic emissions associated with rain drops, and observe that the frequency of the impact radiation is in the hundreds of kHz band. We conclude that impact noise is not a significant source of sound in the breaking region at frequencies below 100 kHz, although the radiation from entrained air may be.

Turbulence acts as a quadrupole acoustic source,[42] and is thus a very inefficient radiator of sound. However, as pointed out by Dowling and Ffowcs Williams,[43] the presence of bubbles in a turbulent field amplifies the radiated sound considerably, even at frequencies well below bubble resonance. Thus the turbulent pressure field beneath the breaking wave in concert with the entrained air may be an important source of sound at lower frequencies. An estimate of the acoustic intensity that can be expected from this mechanism has been made by Prosperetti,[44] but it is not clear that the numerical values he adopts are relevant to bubble plumes in the surf zone. It is not possible to make quantitative statements about the contribution of this mechanism without measurements of the turbulent pressure spectrum in bubble clouds below surf zone waves.

Infrasonic sources of pseudo-sound are nonacoustic (i.e., nonradiating) pressure fluctuations associated with the flow of water around a hydrophone. An analysis of these sources has been given by Strasberg.[31] The significant mechanism in the presence of existing turbulence (i.e., turbulence other than that caused by the presence of the hydrophone) is convection of the turbulent pressure field past the hydrophone. Strasberg presents estimated pressure fluctuations for various flow speeds and frequencies, based on a wave number spectrum for the turbulent velocity field measured off the Canadian Pacific coast at a depth of 89 m by Grant *et al.*[45] From Strasberg's estimates and the measured relative flow between the hydrophone and water of 2.5 m/s during the passage of a breaking wave in our acoustic experiment, we can expect the acoustic measurements made outside the bubble cloud to be free from flow-induced noise above 10 Hz. Measurements by Osborn *et al.*[46] of the open ocean bubble layer and bubble clouds indicate that the turbulence within bubbly regions may be as much as 100 times greater than between and below the regions. In this case, a hydrophone passing through a bubble cloud may measure significant flow noise at low frequencies.

Collective oscillations as a low-frequency source of sound in the ocean were postulated independently by Properetti[47] and Carey[48] who suggested that a cloud of bubbles can interact in unison, creating a resonant system with the dimensions of the cloud and a resonant frequency much lower than that of any of the individual bubbles. Low-frequency emissions from spherical bubble plumes created by dropping a mass of water onto a still-water surface have been reported in a series of studies by Kolaini and colleagues.[49–51] Bubble plumes created in this way undergo volume pulsations and radiate low-frequency sound.[51] A detailed analysis of the behavior of spherical bubble clouds in a pure fluid has been presented by d'Agostino and Brennen.[52] A sufficiently accurate expression for our present purposes for the resonance frequency of a spherical bubble cloud is given by Carey *et al.*,[23]

$$f_0 = \frac{1}{2\pi R}\left(\frac{3p_0}{\rho\alpha}\right), \tag{2}$$

where R is the radius of the bubble plume, p_0 is the equilibrium pressure inside the bubbles, ρ is the density of sea water, and α is the average void fraction of the cloud. For the bubble clouds studied here, R is approximately 0.75 m, $\alpha = 0.34$, $\rho = 1000$ kg/m^3, and $p_0 = 101$ kPa, implying a resonance frequency of around 6 Hz. Even much smaller plumes, order 10 cm, will not contribute energy above roughly 100 Hz.

The mechanism remaining to be discussed is individual bubble oscillations. None of the mechanisms considered thus far are expected to be significant sources of sound above about 100 Hz and below 50 kHz, with the exception of bubble-amplified turbulence which is difficult to quantify without measurements of the turbulent velocity field. There is strong experimental evidence showing that individual bubble resonances are the dominant source of sound from gently spilling breakers in the laboratory[8,53,54] at frequencies approximately greater than 500 Hz. In addition, Loewen and Melville[5] have demonstrated that the sound of gently spilling waves can be explained with an acoustic dipole model of bubbles entrained at the sea surface. Given the preponderance of bubbles entrained by breaking waves in the surf zone and the absence of other significant mechanisms, we hypothesize that individual bubble resonances are the primary source of sound between 500 Hz and 50 kHz for surf zone waves breaking above fine to medium grained sand.

Our conclusions regarding source mechanisms are in substantial agreement with those of Didenkulov and Korotkov[55] for frequencies above 500 Hz, although we have identified sediment disturbance as contributing at substantially higher frequencies than those authors for waves breaking over fine to medium grained sands.

IV. THE SOUND OF AIR ENTRAINMENT

In a consideration of bubble noise mechanisms, Longuet-Higgins[56] reviewed the relationship between bubble formation and bubble oscillations. A bubble emits a pulse of sound when it is formed, an event Longuet-Higgins describes as ''a kind of birth yell.'' Although some of the mechanisms leading to bubble formation at the ocean's surface are understood,[56] the subject remains an area of active investigation. According to Melville,[2] after entrainment a volume of air is broken up into smaller and smaller bubbles. In addition to this process, drops impacting the surface,[37] and penetration of the initial jet[57,58] formed by plunging breakers lead to bubble formation. The processes of bubble formation in the surf zone, where even the gross hydrodynamic features of breaking in the transition region are not yet well understood,[59] are undoubtedly complicated and we will not address them here. Instead we will take an inverse approach, and infer what we can about the nature of the air entrainment process from the acoustic and photographic experiments reported in Secs. I and II.

The photographic observations presented in Sec. II show quite clearly that the bubbles entrained in the transition region are packed into cells, which themselves show a propensity to form lines behind the breaking wave. The existence of bubble plumes obviously implies some hydrodynamic creation process, which we will refer to as a ''plume injection event,'' or simply an ''injection event.'' An injection event is, by definition, the result of bubble formation mechanisms

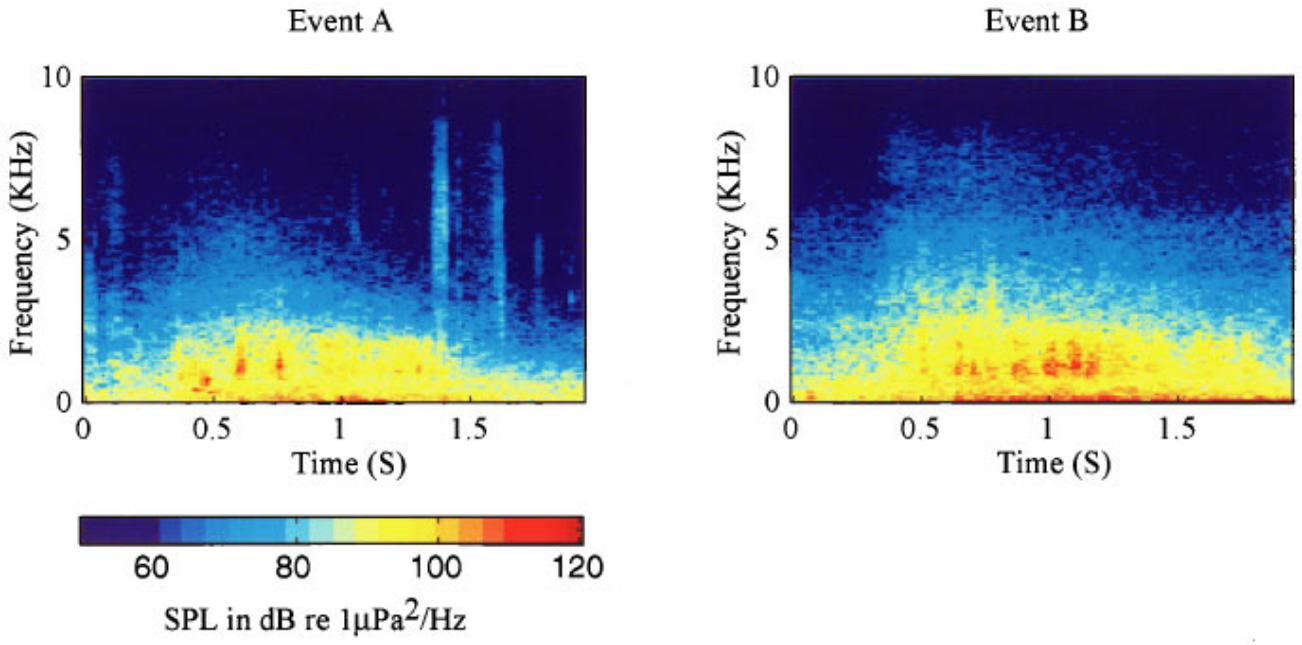


FIG. 8. Frequency versus time plots showing the bursts of acoustic noise generated by two of the smaller waves in the data set. (a) A particularly low-energy wave which broke in a limited region directly over the seaward hydrophone. (b) A more energetic wave showing more noise bursts.

coupled with those hydrodynamic forces which result in the distribution of the bubbles into a plume.

Accepting that the dominant source of sound above 500 Hz results from the creation of bubbles, a question of importance to the wave noise arises. Are the bubbles formed simultaneously with the bubble plume, or is the creation process continuous with a subsequent reorganization into plumes? Bubble creation simultaneous with plume formation would result in bursts of noise, whereas continuous creation would result in a steady background noise. There is evidence from the acoustic data that the wave noise, and therefore bubble formation, does in fact occur in bursts.

Figure 8(a) (event A) and (b) (event B) shows spectrograms of the active period for two class 2 events. Event A was a particularly low-energy wave which broke in a limited region directly above the seaward hydrophone. Event B was more energetic, and broke over a greater lateral extent than A. Both spectrograms show a number of vertical bands corresponding to broadband events roughly 50 ms in duration. The bursts of acoustic energy are direct evidence that the bubble formation process is intermittent in nature. Although this does not prove they are associated with the formation of a bubble plume, it is supporting evidence for the hypothesis.

Pulling the photographic and acoustic evidence together, a consistent picture of the sound source begins to emerge. As a wave steepens and breaks, a series of bubble plume injections occur. Each injection results in the formation of bubbles and a bubble plume which radiates a burst of acoustic energy into the water column lasting approximately 50 ms. For waves with injection rates of more than about 20 events per second, individual acoustic bursts will overlap and the wave spectrum will appear relatively uniform. Waves with a lower injection rate will show the vertical bands evident in Fig. 8(a) and (b). Collective oscillations, bubble-amplified turbulence, and surfseisms are all mechanisms which may generate sound below 100 Hz. Between 100 and 500 Hz there will be a number of contributing mechanisms—the low-frequency sources discussed above which are decreasing in level but still significant, and the sound of resonating bubbles which are rising in level. The noise of drop impacts and disturbed sediment contribute to the noise spectrum above about 50 kHz, and may be discernible at frequencies as high as 400 kHz.

V. A SOURCE MODEL

Our model for the sound of a breaking wave in the surf zone is based on the hypothesis that ringing bubbles dominate the spectrum between 500 Hz and 50 kHz. A consistent interpretation of the acoustic and photographic measurements indicates that active bubbles are created during a bubble plume injection event. Our analysis begins, then, with the sound of a plume of ringing bubbles below the surface of a wave guide. The dynamics of bubbly flows have been studied extensively in the last few decades (see, for example, d'Agostino and Brennen[52]). It is beyond the scope of the present work to present a detailed analysis of an acoustically active bubble plume in a shallow water wave guide. In what follows we will attempt to construct a simple source model, introducing approximations where necessary.

Consider a bubble plume created in a pure fluid as shown in Fig. 9. The plume is composed of bubbles of various radii a, with a size distribution $n(a)$ bubbles per cubic meter, per unit radius increment, scattered throughout the plume volume. Assuming that the bubbly mixture can be modeled as a linear, time-invariant system for the time intervals required to interpret the temporal evolution of the breaking wave spectrum, the acoustic pressure observed at a point $\mathbf{O}$ due to a single bubble created at time t_0 within the volume element $d\mathbf{r}$ is

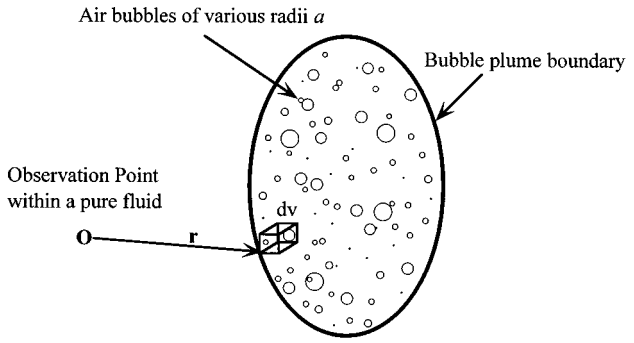


FIG. 9. The geometry for the bubble plume noise model. Bubbles of various radii are assumed to be distributed throughout the plume which is surrounded by a pure fluid.

$$b(t)=\int_{-\infty}^{t} s(\tau-t_0,a)g(t-\tau,\mathbf{r})d\tau, \tag{3}$$

where $\mathbf{r}$ is the distance from $\mathbf{O}$ to the volume element dv, $s(t,a)$ is the pressure pulse radiated by the bubble, and $g(t,\mathbf{r})$ is the impulse response of the bubbly mixture within the plume and the fluid surrounding the plume.

Suppose that within the small element of volume dv, and during the period that the plume is acoustically active τ_c, there are $\nu(a,\mathbf{r})da\ dv$ bubbles with radius between a and $a+da$ created at various times t_j which radiate sound. Assuming that the radiation of neighboring bubbles has a negligible effect on an individual bubble pressure pulse, the contribution to the acoustic field at $\mathbf{O}$ from the radiating bubbles within dv is given by the sum of the individual contributions:

$$dp(t)=\sum_{j=1}^{\nu(a,\mathbf{r})da\ dv}\ \int_{-\infty}^{t} s(\tau-t_j,a)g(t-\tau,\mathbf{r})d\tau. \tag{4}$$

If the bubble creation times t_j are uniformly randomly distributed during the acoustically active period of the plume, the bubble pulses s are Poisson distributed. In this case, the power spectrum of $dp(t)$ is a known function, and is given by

$$dP(\omega)=\lambda|S(\omega,a)G(\omega,\mathbf{r})|^2\ da\ dv, \tag{5}$$

where $\lambda(a,\mathbf{r})=\nu(a,\mathbf{r})/\tau_c$ is the bubble creation rate as a function of bubble radius and position, and S and G respectively are the Fourier transforms of s and g. The observation that the power spectrum of an ensemble of bubbles depends on the bubble creation *rate* has also been noted by Crowther.[60] Equation (5) is an expression for the contribution to the sound intensity at $\mathbf{O}$ due to the active bubbles with radii between a and $a+da$ in the volume element dv. Before integrating Eq. (5) over radius and volume to obtain the total power spectrum, we will consider the propagation of sound radiated by bubbles at different positions within the plume by studying the acoustic properties of the bubbly mixture inside the plume.

A ringing bubble within a plume is in a very different environment than a bubble outside it. The high void fraction in the plume lowers the speed of sound to a small fraction of its value in the pure fluid, and the bubbles scatter and absorb sound. Clay and Medwin[61] give a dispersion relation for a bubbly mixture which accounts for absorption by the constituent bubbles, which are assumed to be of uniform radius. The equation, in terms of the complex propagation constant of a plane wave radiating through the material, is

$$k(\omega)=\frac{\omega}{c_0}\left(1+((\omega_r/\omega)^2-1-i\delta)\times\frac{4\pi aNc_0^2/\omega^2}{((\omega_r/\omega)^2-1)^2+\delta^2}\right)^{1/2}, \tag{6}$$

where c_0 is the speed of sound in the pure fluid, N is the number of bubbles per unit volume, a is the bubble radius, ω_r is the bubble resonance frequency in units of rads/s, and δ is the bubble damping constant representing thermal, acoustic radiation, and viscous losses. Equation (6) provides a more general description for propagation through the bubbly mixture than Wood's equation[62] in that it accounts for dispersion of sound, and acoustic absorption by the bubbles within the mixture.

An expression for the bubble resonance frequency assuming adiabatic oscillations and neglecting the effects of surface tension (which is sufficiently accurate for our purposes) is[63]

$$\omega_r=\frac{1}{a}\left(\frac{3\gamma P}{\rho}\right)^{1/2}, \tag{7}$$

where γ is the ratio of specific heats of the gas, P is the quiescent pressure inside the bubble and ρ is the density of the pure fluid (values of γ used for air vary between 1.4 and 1.6, and $\rho=1.03\times10^3$ kg/m^3 for sea water). An expression for the damping constant as a function of bubble radius can be found in Clay and Medwin.[61]

As described by Clay and Medwin,[61] Eq. (6) can be extended to account for a bubble size distribution by replacing the number of bubbles N by the factor $n(a)da$, where $n(a)$ is the bubble size distribution as a function of radius, and integrating the resulting expression over bubble radius. The resulting expression is

$$k(\omega)=\frac{\omega}{c_0}\left(\int_{a_0}^{a_1}1+((\omega_r/\omega)^2-1-i\delta)\times\frac{4\pi ac_0^2n(a)/\omega^2}{((\omega_r/\omega)^2-1)^2+\delta^2}\,da\right)^{1/2}, \tag{8}$$

where a_0 and a_1 are the lower and upper limits on the bubble size distribution radii. A detailed derivation of Eq. (8) has been presented by Commander and Prosperetti,[64] who have compared it with data sets of other investigators. They found that the model works well up to volume fractions of 1%–2%, provided that bubble resonances play a negligible role. The void fractions of the bubble plumes considered here are significantly greater than 2%, and we can only consider the application of Eq. (8) as indicative of their properties.

A plane wave propagating through the bubbly mixture has a phase speed of $c=\Re(\omega/k)$ and is attenuated at the rate $\beta=-\Im(k)$ nepers per meter of propagation path. Equation (8) has been evaluated as a function of frequency using c_0

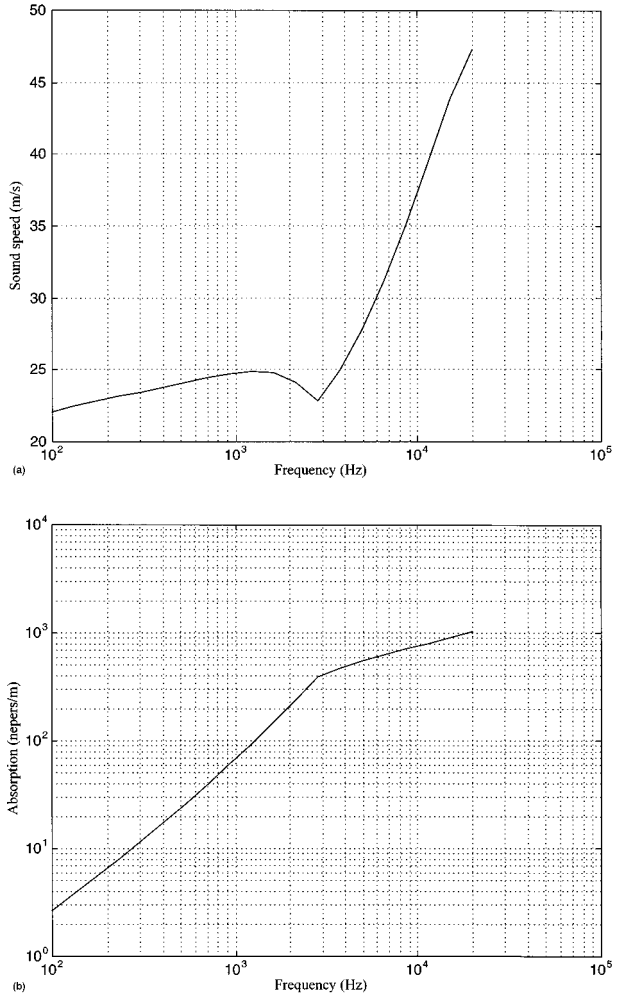


FIG. 10. Values of (a) the phase speed and (b) the absorption of sound in a bubble plume calculated from Eq. (8) and plotted as a function of frequency.

$=1500$ m/s, Eq. (7) for ω_r, and Clay and Medwins' expression for δ. The bubble size distribution used was chosen as a fit to the photographic measurements presented in Sec. II B, and was taken to be

$$n(a)=\begin{cases} n_0\left(\dfrac{a}{a_0}\right)^{-2.5}, & 100\ \mu\text{m}\leqslant a<1.3\ \text{mm}, \\ n_0\left(\dfrac{a}{a_0}\right)^{-4.5}, & 1.3\ \text{mm}\leqslant a<5\ \text{mm}, \end{cases} \tag{9}$$

where $n_0=10^4$ bubbles per m^3 per unit radius increment and $a_0=1.3$ mm. The results of the calculation are presented in Fig. 10(a) and (b) as values for the phase speed c and the absorption β for frequencies between 100 Hz and 20 kHz. Beyond 20 kHz, propagation through the plume has an increasing sensitivity to the presence of bubbles smaller than 100 μm. As the bubble size distribution rolls off for radii less than 100 μm, we have not considered frequencies greater than 20 kHz.

For frequencies less than roughly 3 kHz, the sound speed is constant at just over 20 m/s, a result which agrees favorably with sound speed estimates made by Loewen and Melville[6] inside fresh water bubble plumes entrained by laboratory waves. For frequencies greater than 3 kHz, the

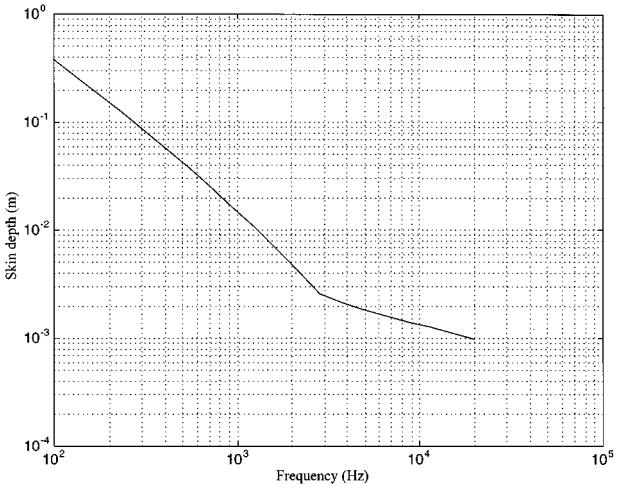


FIG. 11. The skin depth of a bubble plume calculated from the absorption values shown in Fig. 10(b) and plotted as a function of frequency. Bubbles radiating sound more than a skin depth or so within the plume are inaudible from the plume exterior.

bubbly mixture becomes dispersive and the phase velocity increases to about 50 m/s at 20 kHz. The absorption by the plume increases from 3 nepers per meter at 100 Hz to 1000 nepers per meter at 20 kHz. These absorption values are high and result in the rapid attenuation of acoustic radiation. We cannot compare our results with Loewen and Melville[6] as they did not present estimates of the plume absorption. The change in behavior of the sound speed and absorption at 3 kHz is a consequence of the change in the scaling law for the number density at bubble radius $a=1.3$ mm.

A plane wave of angular frequency ω propagating through the bubbly mixture attenuates by a factor of $1/e$ every $1/\beta$ m, an observation which leads us to introduce the idea of a frequency-dependent ''skin depth'' $L_a(\omega)=1/\beta(\omega)$. The energy from bubbles radiating more than a skin depth or so away from the observation point is rapidly absorbed by the bubble plume. The skin depth corresponding to the β curve of Fig. 10(b) is plotted in Fig. 11. The depth varies from around 40 cm at 100 Hz to 1 mm at 20 kHz, implying that only the active bubbles in a relatively thin shell on the plume boundary contribute to the sound radiated.

The bulk of the energy radiated by bubbles within the plume is absorbed before reaching the plume boundary; only bubbles within a skin depth or so of the boundary are audible beyond it. Accordingly, we have modeled the audible region of the plume as a sheet of bubbles with the same height and length as the plume, and width given by the skin depth. As the skin depth is a decreasing function of frequency, the effective volume of the bubble sheet also decreases with frequency. A detailed model would account for the varying sound speed and rough boundary at the plume's edge. Lacking measurements of the internal structure of the plumes, we have made the simplifying assumption that the sound radiated by a bubble in the sheet is unaffected by the presence of the sheet.

Our simplified model for the sound of a plume consists of a stationary sheet of ringing bubbles below the surface of a shallow water wave guide (see Fig. 12). The power spec-

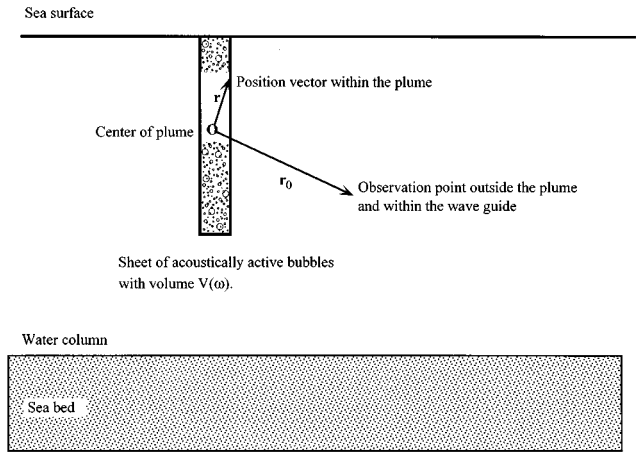


FIG. 12. The geometry for the ''bubble sheet'' model of sound radiated by an acoustically active bubble plume.

trum at a point $\mathbf{r}_0$ outside a plume is found by integrating Eq. (5) over the sheet volume $V(\omega)$, which depends on the skin depth, and is therefore a function of frequency, and radius of the bubble size distribution within the sheet:

$$P(\omega,\mathbf{r}_0)=\int_{a_0}^{a_1}\int\int_{V(\omega)}\int \lambda(a,\mathbf{r})|S(\omega,a) \times G(\omega,\mathbf{r},\mathbf{r}_0)|^2\, d\mathbf{r}\, da, \tag{10}$$

where $\lambda(a,\mathbf{r})$ is the bubble creation rate as a function of bubble radius and position within the sheet, which has units of number of bubbles created per cubic meter of active volume per unit bubble radius increment per second, and a_0 and a_1 are the minimum and maximum length scales of the bubble size distribution within the sheet. There are few measurements of the creation rate of bubbles; Longuet-Higgins has given a recent review.[56] The measurements which do exist are of the rate of bubble entrainment per unit area of the sea surface per unit time, which seem to make the tacit assumption that bubbles are not created within the water volume. Bubble fragmentation and coalescence are two bubble formation mechanisms that may be important in the water column within bubble plumes.

Because we are assuming that the radiation from bubbles is unaffected by the sheet, $G(\omega,\mathbf{r},\mathbf{r}_0)$ in Eq. (5) is the Greens function for propagation in the shallow water wave guide as a function of angular frequency ω, position within the bubble plume $\mathbf{r}$ and the observation point relative to the center of the sheet $\mathbf{r}_0$. An expression for the bubble spectrum S can be found in Loewen and Melville,[5] although we do not expect the acoustic dipole model of bubbles entrained at the sea surface adopted by them to explain the sound of gently spilling waves to be applicable to the energetic breaking events reported here.

Equation (10) is our final expression for the power spectrum of the noise generated by the creation of a bubble plume. It applies while there are bubbles being created on the boundary of the plume; the acoustic measurements indicate this time is of order 50 ms. The sound of the breaking wave between 500 Hz and 50 kHz is formed from the superposition of multiple plume injection events, each of which has a spectrum of the form given by Eq. (10). As we discussed earlier, if there are more than roughly 20 plume injections per second, the bursts of sound which accompany the injections overlap, and the resulting noise is continuous. Thus for sufficiently energetic waves (assuming that the injection rate scales according to wave energy), we can model the wave noise by a bubble sheet that extends along the entire front of the wave, and moves through the water column at the forward speed of the breaking region. Equation (10) gives an expression for the power spectrum of the advancing sheet of bubbles, provided the volume of the space integral is taken to be across the entire front of the breaking wave. We will use this simple model in the next section when we consider propagation of wave noise through the surf zone. A more sophisticated treatment would follow the same lines as Finette and Heitmeyers' [65] analysis of the open ocean noise field and model the formation of individual plumes statistically.

VI. PROPAGATION THROUGH THE SURF ZONE

There are two main environmental features that characterize acoustic propagation in the surf zone: the proximity of the sea floor and the presence of bubbles entrained by the relentless breaking of waves.

The bathymetry of beaches is complicated, and depends on, among other factors, the grain size of the beach sediment, exposure of the beach to wave attack, and the tidal range.[66] In addition, almost all beaches are subject to frequent large fluctuations in size and shape.[66] Onshore sediment migration occurs during long-period waves, while large waves of short period cause erosion. The result of this action are winter and summer cycles caused by the seasonal presence of storms. These processes are obviously not a significant factor for our experiment, which lasted for less than a day, but will affect the propagation of sound from the surf zone over seasonal time scales.

For horizontal scales greater than roughly 100 m, changes in the water depth due to the gross slope of the sea floor affect the sound field and need to be included in a model for propagation. Despite variable bathymetry, and notwithstanding the possibility of longshore bars and troughs, the simplest geometrical representation for a beach is a wedge. The literature dealing with propagation in wedges is extensive and the reader is referred to Buckingham[67] for a discussion. Recent developments include three-dimensional solutions for propagation in a wedge over a penetrable sea floor.[68–70] In the present analysis, we will restrict attention to short length scales (order 10 m) over which changes in the water depth due to gross sea floor slope are unimportant in calculations of the sound field.

The entrainment of air by breaking surf is a very evident process. Degasing bubble clouds can be seen (and heard in air) for tens of seconds after the passage of the wave which created them. The persistence of bubbles in the water column results in a residue of air which absorbs and scatters sound from the shoreline to the seaward edge of the transition region. Plumes are initially created in a reasonably ordered pattern, as shown in Fig. 4. In the deep ocean, bubble plumes

are driven by the forces of buoyancy, turbulent diffusion, and dissolution, and evolve in a predictable way.[21] In the surf zone, the presence of rips and undertows and the repeated passage of breaking waves over plumes makes their evolution more difficult to model. Qualitative observations from the surface show that the spatial distribution of the bubbles is complicated and time dependent.

Despite an environment complicated by absorption due to acoustically quiescent bubble plumes, and scattering due to surface roughness, we will use a simple environment to model propagation through the wave guide. The environment is shown in Fig. 12, and neglects both bubble plume absorption and surface scattering. The justification for this neglect is not that these effects are, *a priori*, expected to be unimportant, but that we can use the differences observed between the modeled and observed propagation to provide a measure of them. If the modeled and observed sound fields compare favorably, then we can conclude that the bubble residue, and surface and bottom roughness scattering do not dominate propagation over the length scale (order 10 m) and frequency range we are considering (400–5000 Hz).

A. The spectral ratio

The purpose of this section is to introduce a measurement which allows the separation of source spectrum and propagation effects on the power spectrum of a wave observed at some distance from it. The basis of this measure, which we will call the ''spectral ratio,'' comes from a qualitative observation of the bubble plumes obtained from the photographic study. For reasons described in Sec. II, the photographic images of bubble plumes were taken at random depths within the first 0.5 m of the water column. The images showed a relatively homogeneous distribution of bubbles of different sizes in the vertical, both within a given image, and from one image to another. The homogeneous distribution of bubbles observed at random depths indicates the size distribution is independent of depth within 0.5 m of the surface. Such a condition is unlikely to be true for evolved bubble plumes, where larger bubbles have been driven to the surface by buoyancy, and small bubbles driven down by turbulent diffusion. However, the photographic evidence suggests a uniform spatial distribution of different sized bubbles within approximately a second after the instant of creation. In this case, the bubble creation rate can be expressed as a product of a number density function and a space function:

$$\lambda(a,\mathbf{r}) = n_r(a) f(\mathbf{r}), \tag{11}$$

where a is the bubble radius and $\mathbf{r}$ is position within a bubble plume. The function $n_r(a)$ is the size distribution creation *rate* of the acoustically active bubbles, is assumed to be constant throughout the creation of the plume, and has units of bubbles per unit volume, per unit radius increment, per unit time. The space function, $f(\mathbf{r})$, is a dimensionless function relating the magnitude of the size distribution creation rate at different points in space. Clearly there is an arbitrariness to $f(\mathbf{r})$ unless we choose an additional constraint, so we set $f(0)=1$.

Substitution of Eq. (11) into Eq. (10) results in an expression for the power spectrum of the breaking wave which is a product of two integrals:

$$P(\omega,\mathbf{r}_0) = \int_{a_0}^{a_1} n_r(a) |S(\omega,a)|^2 \, da \times \int\int\int_V f(\mathbf{r}) |G(\omega,\mathbf{r},\mathbf{r}_0)|^2 \, d\mathbf{r}. \tag{12}$$

The integral of the number-density-weighted bubble-pulse spectra over radius is independent of the observation point $\mathbf{r}_0$. This factor represents the source power spectrum, and we will subsequently refer to it as the source integral. The integral over $\mathbf{r}$ is independent of number density, and represents the effect of the spatial distribution of active bubbles on the power spectrum observed at $\mathbf{r}_0$. We will refer to this integral as the Green's function integral. As we noted earlier, the sheet volume V is determined by the plume skin depth, and is strictly a function of frequency. We have not included this frequency dependence in Eq. (12) for the following reason. The skin depth at frequencies above 5 kHz is less than 2 mm, which is considerably less than roughness scale associated with the plume boundary [see Fig. 4(b)]. Keeping the sheet thickness constant will overestimate the volume of the active region at higher frequencies, but will be more representative of the length scale over which the active bubbles are distributed within the outer edge of the plume.

As the source integral is independent of position, we can eliminate it from Eq. (12) by taking the ratio of the observed spectrum P at different points in space, $\mathbf{r}_0$ and $\mathbf{r}_1$ say. The resulting ratio, which we will call the spectral ratio,

$$R(\omega,\mathbf{r}_0,\mathbf{r}_1) = \frac{P(\omega,\mathbf{r}_1)}{P(\omega,\mathbf{r}_0)} = \frac{\iiint_V f(\mathbf{r}) |G(\omega,\mathbf{r},\mathbf{r}_1)|^2 \, d\mathbf{r}}{\iiint_V f(\mathbf{r}) |G(\omega,\mathbf{r},\mathbf{r}_0)|^2 \, d\mathbf{r}}, \tag{13}$$

depends on the Green's function integral, but is independent of the number density and spectrum of ringing bubbles in the sheet. According to Eq. (13), the ratio of the acoustic power spectrum measured at two points in the water column is sensitive to propagation through the wave guide but independent of the source spectrum. This property allows source spectrum and propagation effects to be separated, and makes the ratio a suitable basis for comparison between a model for the Green's function and the measured transmission properties of the wave guide.

B. Calculation of the spectral ratio

The geometry for the spectral ratio calculations is shown in Fig. 13. For Eq. (13) to be valid, the spectral ratio must be calculated when the acoustic field is dominated by the noise radiated by a nearby breaking wave. Accordingly, we have chosen a time when a breaking wave lies between the two hydrophones deployed during the acoustic experiment. The wave guide is assumed to be isovelocity and of constant depth. As discussed in Sec. V, the source of acoustic radiation is taken to be a sheet of bubbles moving through the

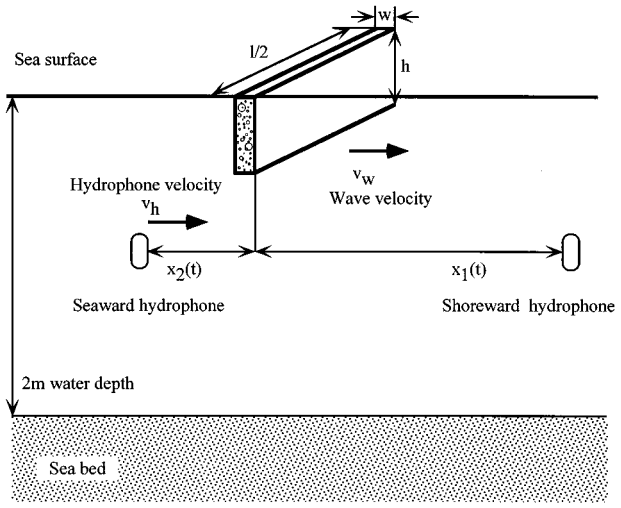


FIG. 13. The geometry for the spectral ratio calculations which characterize propagation in the surf zone wave guide.

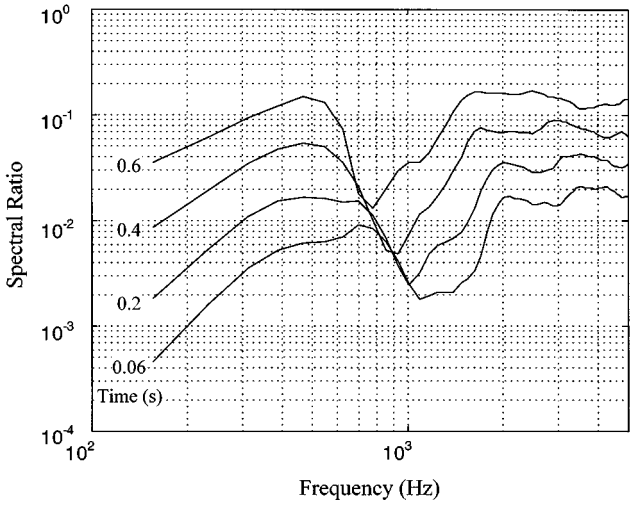


FIG. 14. Theoretical spectral ratios calculated by evaluating the integrals in Eq. (13). The ratio is computed at four different times after the acoustically active bubble sheet has passed by the seaward hydrophone.

water column from left to right with speed ν_w, which passes over the seaward hydrophone at time $t=0$. The shoreward movement of the seaward hydrophone, discussed in Sec. I, has been included in the model.

The spectral ratio was calculated at four times, the first of which corresponds to the passage of the active region above the seaward hydrophone. As time progresses, the active region recedes from the seaward hydrophone and advances toward the shoreward hydrophone. The net result is an increase in sound level at the shoreward hydrophone relative to the seaward hydrophone, and an increase in the spectral ratio with time. The parameters chosen for the ratio calculations were $d=1$ m, $w=0.25$ m, and $l=8$ m for the dimensions of the bubble sheet, $\nu_w=5$ m/s for the sheet speed, and $\nu_h=2.5$ m/s for the seaward hydrophone speed. The bubble creation distribution within the sheet was assumed to be uniform, so that $f(\mathbf{r})=1$ throughout the sheet volume. The water depth and sound speed were taken to be 2 m and 1500 m/s, respectively. The compressional wave speed, density, and absorption of the sea floor were set to $c_p=1700$ m/s, $\alpha=0.5$ dB/wavelength, and $\rho=1500$ kg/m^3, respectively, which are representative of the geoacoustic properties of the sand around the pier.

The greatest uncertainty in the parameters chosen for the sheet is in the estimate of the dimensions of the bubble sheet and the dimensionless function $f(\mathbf{r})$. Calculation of the spectral ratio for a variety of sheet volumes and functional forms for $f(\mathbf{r})$ showed that the ratio is fairly insensitive to these quantities. The parameters listed above were not chosen because they yielded an especially good fit to the data set; they are one of a number of reasonable choices, all of which yield similar results.

Because the seaward hydrophone is in the near field of the radiating bubble sheet, the mathematical formulation used to calculate the Green's function in the wave guide needs to be accurate for ranges less than a few tenths of a wavelength from a source. In this case, and because the canonical geometry we have chosen does not include surface roughness or a sound speed profile, the image formulation discussed by Brekhovskikh[71] (which includes the effects of modal cutoff in the wave guide) yields a satisfactory mathematical solution.

The Green's function integrals in Eq. (13) were calculated numerically using a simple trapezoidal rule, and the parameters listed above at $t=0.06$ s, $t=0.2$ s, $t=0.4$ s, and $t=0.6$ s.

C. Comparison between theory and observation

Figure 14 shows four theoretical spectral ratios corresponding to four times, plotted as a function of frequency and computed as described in Sec. VI B. Figure 15 shows four measured spectral ratios corresponding to the same times as Fig. 14, obtained by averaging nine class 1 events. Energetic events were chosen for the averages to ensure that the sound of the breaking wave was well above the background noise. The time at which the active period passed

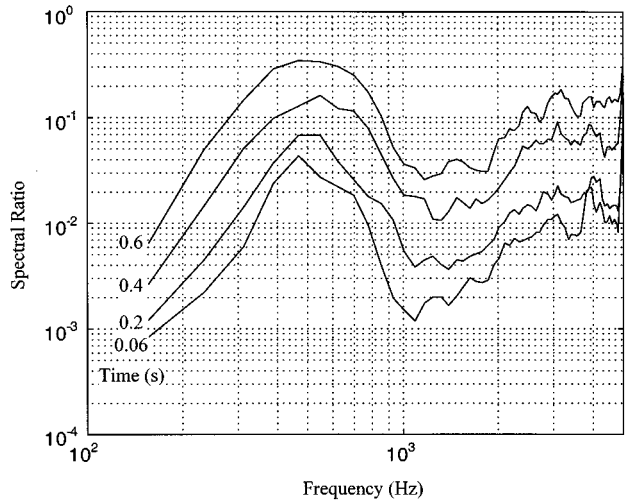


FIG. 15. Experimental spectral ratios estimated from the ratio of the observed power spectra at the seaward and shoreward hydrophones, averaged over 9 class 1 breaking events. The ratio is measured at four times after the active region of the breaking wave has passed over the seaward hydrophone.

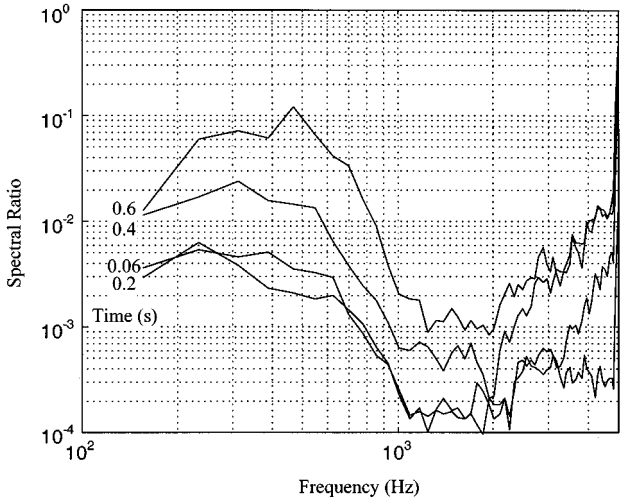


FIG. 16. An example of spectral ratios obtained from a breaking event which did not agree with the moving bubble sheet model. The curves show a strong absorption mechanism not accounted for by the model.

over the seaward hydrophone, corresponding to $t=0$, was identified as the start of the active period in a spectrogram of the breaking event.

The theoretical and measured spectral ratios agree favorably. Both sets of curves show a peak at 500 Hz and a dip around 1.5 kHz. In addition, the trend of the ratio to increase with increasing time are similar, as are the overall ratio levels. The roll-off in the ratios below 500 Hz is due to modal cut off in the wave guide, and the dip at 1.5 kHz is a modal interference effect. There are features of the curves which differ. The measured ratios are consistently smaller than the theoretical ratios between 1 and 3 kHz, indicating an absorption or scattering mechanism present in the surf zone which was not included in the simple environment used to model the acoustics. Two obvious candidate mechanisms are absorption by the bubble plume residue discussed in Sec. VI, and scattering associated with surface deformations.

Two of the nine events averaged did not show good agreement with the theoretical calculations and one of these is shown in Fig. 16. Although the general shape of the measured ratios is similar, with a peak around 400 Hz and a dip at 1.5 kHz, the ratio levels are roughly an order of magnitude smaller than those calculated between 1 and 5 kHz indicating a strong absorption or scattering mechanism.

In addition to the spectral ratio, the power spectrum of the sound field at the seaward hydrophone at different times during the passage of the wave was computed. Figure 17(a) shows the acoustic power spectral density measured at the seaward hydrophone during the passage of a breaking wave. The power spectral density divided by the Green's function integral computed at the appropriate range (labeled as ''compensated spectrum'') is shown in Fig. 17(b), and the Green's function integral is shown in Fig. 17(c). As might be expected, the level of the spectra in Fig. 17(a) decay as time progresses and the range between the hydrophone and the active region of the wave increases. By dividing the spectrum by the Green's function integral, evaluated at different ranges and computed using the canonical model, a significant reduction in the systematic variation in level between the

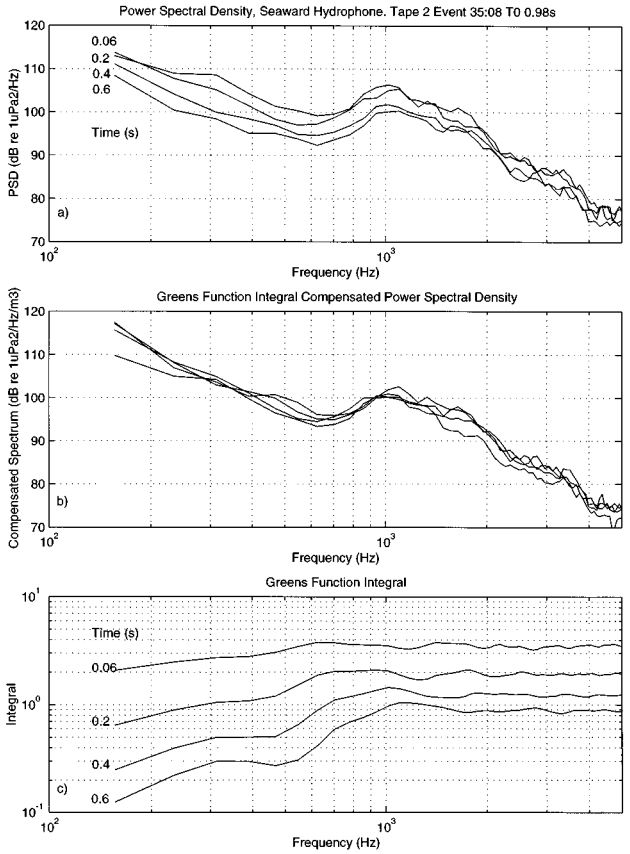


FIG. 17. Power spectral densities measured at four times after the active region of the breaking wave has passed over the seaward hydrophone (a) before and (b) after normalization by the Greens function integral [defined in Eq. (12)]. (c) The Greens function integral plotted at the same four times as the power spectral density.

curves at different times can be seen [see Fig. 17(b)]. That is, the Green's function integral predicts the decay in level due to increasing range with reasonable accuracy. The result, which does not rely on a ratio but an absolute computation of the Green's function integral, and which is frequency dependent, is a strong indication that the canonical model is reproducing the first-order physics of the environment.

We conclude from these comparisons that propagation calculations based on a relatively simple geometry can reproduce the first-order physics of the sound field observed during a breaking wave event in the transition region of the surf zone over length scales of order 10 m and between 200 and 5000 Hz for roughly 80% of the breaking waves.

VII. THE SOURCE SPECTRUM

The spectral ratio calculations demonstrate that a simple environmental model can be used to describe acoustic propagation over short distances in the transition region of the surf zone. Having established a propagation model, it is possible to make an estimate of the source integral, which is effectively a measure of the acoustic spectrum of a breaking wave. Rearrangement of Eq. (12) yields an expression for the source integral in terms of the measured power spectrum and the Green's function integral:

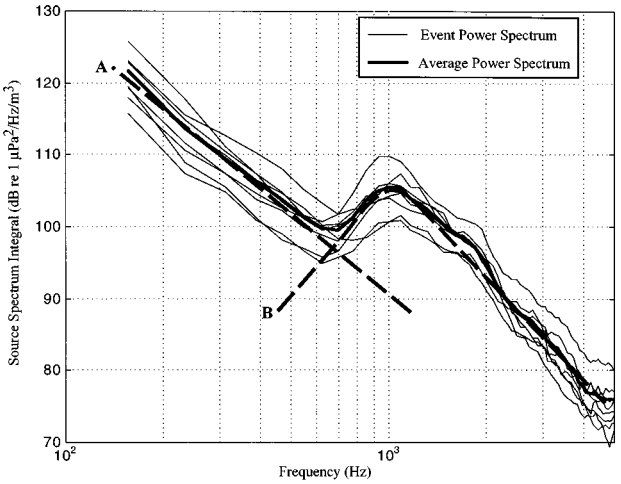


FIG. 18. An estimate of the source integral (which represents the acoustic power radiated by a breaking wave) obtained from Eq. (14). The power spectrum is averaged over 9 class 1 breaking events.

$$\int_{a_0}^{a_1} n(a)|S(\omega,a)|^2\, da = \frac{P(\omega,\mathbf{r}_1)}{\iiint_V f(\mathbf{r})|G(\omega,\mathbf{r},\mathbf{r}_1)|^2\, d\mathbf{r}}. \tag{14}$$

The source integral as a function of frequency estimated from the same nine class 1 events as the spectral ratios are shown in Fig. 18. The units of the integral in the figure are dB referenced to a power of 1 μPa^2 per Hz, per cubic meter of bubble sheet. The average curve plotted in Fig. 17 is an estimate of the acoustic power radiated per cubic meter. The two broken lines labeled A and B are discussed below.

Unlike the spectral ratio, the source integral is sensitive to our estimate of the bubble sheet volume. The curves in Fig. 18 are an average of the acoustic power spectrum over the active period, and so the correct volume to use is the average volume of the acoustically active region of bubble plume, adjusted for the skin depth effect. The scatter in the source integral curves occurs because of the variation of the bubble sheet volume from one event to another. The uncertainty in the estimate of the sheet volume represents the greatest source of error in the source spectrum, and could be significant. Quantitative measurements of the number and volume of bubbles plumes created by a breaking event are required to obtain an accurate estimate of the absolute level of the source integral. However, the shape of the integral should be accurately estimated.

The source integral shows a local maximum at 1 kHz, rising 5 dB between 700 Hz and 1 kHz to a peak value of 105 dB, and diminishing at 50 dB per decade beyond 1 kHz. As the effect of the near-by sea floor and sea surface on the source spectrum was accounted for when we divided by the Green's function integral, the maximum at 1 kHz must be due to a source mechanism. In Fig. 18 we have broken the average source spectrum into two curves, labeled A and B. Curve A rolls off at a constant 40 dB per decade, and dominates the spectrum at frequencies less than 700 Hz. Curve B increases to 1 kHz where it reaches a maximum and then decreases, and dominates the spectrum at frequencies greater than 700 Hz. Curve B is very similar in shape to the acoustic power spectrum from gently spilling laboratory waves measured by Medwin and Daniel.[8] Their measured power spectrum, averaged over 6 breaking events, has a peak value of just over 80 dB at 1 kHz and diminishes at 25 dB per decade beyond 1 kHz. Work by Medwin and Daniel,[8] and Loewen and Melville[5] provides strong evidence that the acoustic spectrum from gently spilling waves is due to ringing bubbles. At the end of Sec. III we postulated that individual bubble resonances are the primary source of sound between 500 Hz and 50 kHz for surf zone waves breaking above fine to medium grained sand. The similarities between Medwin and Daniels' measurements and our own in this frequency range suggest that curve B in Fig. 18 is indeed due to ringing bubbles.

VIII. CONCLUDING REMARKS

Plumes of bubbles are formed beneath breakers in the surf zone. The injection of air begins when the wave first breaks and continues through the transition region with later injection events occurring nearer the shore. Once created, a plume persists for tens of seconds and a breaker thus creates a ''wall'' of bubbles which begins where the wave breaks and grows in width as the breaking region propagates shoreward. Individual bubble plumes can be as large as 0.5 m^3 below 1-m waves, have a propensity to be packed into lines behind the breaking wave and have an average void fraction of roughly 0.3–0.4.

Spectrograms of the noise radiated by a breaking wave show an acoustically active region of water which propagates with the wave, located between the leading edge and wave crest. The evidence from the acoustic and optical measurements is that the primary source of sound from the active region between 500 Hz and 5 kHz are the pressure pulses which accompany the formation of bubbles. The source of low-frequency sound below 500 Hz has not been determined, although a review of candidate mechanisms suggests bubble-amplified turbulence and the collective oscillations of bubbles within a plume could play a role. Measurement of the turbulent velocity field inside and nearby a bubble plume will help establish the source mechanism(s) below 500 Hz.

Small breakers show bursts of sound in their spectrograms. The episodic nature of sound production for these waves is consistent with the hypothesis that the wave noise and bubble creation occurs simultaneously with the formation of bubble plumes. The acoustic and optical evidence that sound production is associated with the formation of bubble plumes leads to the interesting speculation that the air entrainment and bubble formation processes could be investigated by a study of the noise they make.

High average void fractions have been estimated for the bubble plumes beneath a breaker, leading to the conclusion that only bubbles on a thin shell of the bubble plume are audible outside it. Photographic measurements of the bubble size distribution within a plume show a decline in bubble count below 100 μm, a result roughly consistent with Medwin and Daniels' acoustical measurement of bubble production by spilling laboratory breakers, although the possibility that we are seeing the size resolution of the optical technique cannot be discounted. The density of bubbles scales as $a^{-2.5}$

for small bubbles and $a^{-4.5}$ for large bubbles, with the transition in scaling law occurring around 1 mm radius. This scaling implies that the void fraction of air is dominated by bubbles with radius greater than 1 mm.

The favorable agreement between the observed and modeled ratio of acoustic power spectra at two horizontally separated points suggest that a simple wave guide calculation is sufficient to account for the first-order physics of acoustic propagation over ranges less than 10 m in the surf zone between 200 and 5000 Hz. Two of the breaking events studied showed the presence of an absorption or scattering mechanism not accounted for by the theory. Two candidate mechanisms are absorption by bubble plumes, and scattering due to surface deformations. It is anticipated that propagation over longer ranges will be influenced by bathymetric changes which could be modeled to first order as a penetrable wedge.

ACKNOWLEDGMENTS

I am grateful to Ms. Dana Lane and Mr. Mark Brentnall who assisted with the data collection, and Professor Michael Buckingham and Dr. John Penrose for a number of helpful discussions. I am also pleased to acknowledge the work of Dr. Dale Stokes who took the underwater photographs, and Mr. Fernando Simonet who undertook the tedious task of measuring the cross-sectional areas of more than 7500 bubbles. This work was supported by the U.S. Office of Naval Research, Contract Nos. N00014-96-1-0120 and N00014-93-D0141.

APPENDIX. THE EFFECT OF THE PIER ON THE ACOUSTIC MEASUREMENTS

A careful examination of the acoustic data in conjunction with the video recordings has allowed the effect of the pier on the noise spectrum to be identified and quantified. There are two effects: a jet of water from the seaward side of the piling which results in a splash, and a bubbly wake on the shoreward side.

A rising jet of water occurs when a wave hits the seaward side of the pier pilings. A series of these jets occur as a wave moves from piling to piling on its shoreward journey. The jet arcs through the air for several seconds and then causes a splash as it strikes the water surface. The acoustic events labeled ''PS'' in Fig. 3(a) and (b) are the sound of jet splashes. The timing of the jet and the wave passage usually resulted in the splashes occurring outside of the active period, which allowed them to be identified in the spectrograms and avoided in the data analysis.

A bubbly wake occurs on the shoreward side of a piling as a wave propagates by, making the shoreward hydrophone most sensitive to this source of signal contamination. The timing of the wake and the wave passage was such that the wake noise and the beginning of the active period for the shoreward hydrophone coincided. As we did not use data from the active period of this hydrophone, it does not affect our results. The seaward hydrophone was less sensitive to wake noise as it was 6.6 m from the nearest seaward piling (the shoreward hydrophone was 3.4 m from the nearest seaward piling).

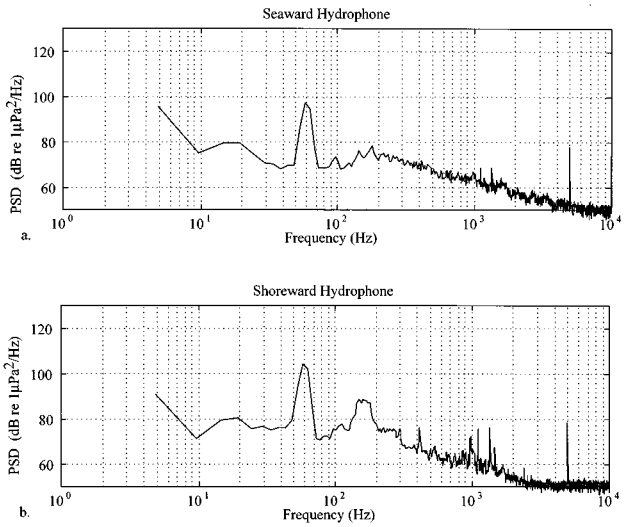


FIG. A1. The background noise power spectrum for a quiet time between breaking waves with the Scripp's Pier pump turned on.

An additional source of noise was the sound of a pump which supplies sea water to Scripps' Aquarium, and which turned on periodically throughout the experiment. The noise power spectral density for a period between waves with the pump on is shown in Fig. A1. The pump spectrum is sharply peaked at 60 Hz, and appears as a 95-dB tone on the seaward hydrophone and 105 dB on the shoreward hydrophone. The two tones just above 1 kHz are also associated with the pump. The 5-kHz tone is an instrumentation artifact. The worst source of contamination comes from the 60-Hz tones which are still at least 10 dB below the levels observed during the active period of a breaking event.

[1] R. J. Urick, *Ambient Noise in the Sea* (Peninsula, Los Altos, CA, 1986).
[2] W. K. Melville, ''The role of surface-wave breaking in air-sea interaction,'' Annu. Rev. Fluid Mech. **28**, 279–321 (1996).
[3] E. Lamarre and W. K. Melville, ''Instrumentation for the measurement of void-fraction in breaking waves: laboratory and field results,'' IEEE J. Ocean Eng. **17**, 204–215 (1992).
[4] E. Lamarre and W. K. Melville, ''Void fraction measurements and sound-speed fields in bubble plumes generated by breaking waves,'' J. Acoust. Soc. Am. **95**, 1317–1328 (1994).
[5] M. R. Loewen and W. K. Melville, ''A model of the sound generated by breaking waves,'' J. Acoust. Soc. Am. **90**, 2075–2080 (1991).
[6] M. R. Loewen and W. K. Melville, ''An experimental investigation of the collective oscillations of bubble plumes entrained by breaking waves,'' J. Acoust. Soc. Am. **95**, 1329–1343 (1994).
[7] W. K. Melville, M. R. Loewen, and E. Lamarre, ''Bubbles, noise and breaking waves: A review of laboratory experiments,'' in *Sea Surface Sound* (Kluwer Academic, Dordrecht, 1993), pp. 483–501.
[8] H. Medwin and A. C. Daniel, Jr., ''Acoustical measurements of bubble production by spilling breakers,'' J. Acoust. Soc. Am. **88**, 408–412 (1990).
[9] P. Papanicolaou and F. Raichlen, ''Wave and bubble characteristics in the surf zone,'' in *Sea Surface Sound: Natural Mechanisms of Surface Generated Noise in the Ocean* (Kluwer, Dordrecht, 1988), pp. 97–109.
[10] V. O. Knudsen, R. S. Alford, and J. W. Emling, ''Underwater ambient noise,'' Marine Res. **7**, 410–429 (1948).
[11] B. R. Kerman, *Sea Surface Sound: Natural Mechanisms of Surface Generated Noise in the Ocean* (Kluwer, Dordrecht, 1988).
[12] B. R. Kerman, *Natural Physical Sources of Underwater Sound: Sea Surface Sound (2)* (Kluwer, Dordrecht, 1993).
[13] O. B. Wilson, S. N. Wolf, and F. Ingenito, ''Measurements of acoustic ambient noise in shallow water due to breaking surf,'' J. Acoust. Soc. Am. **78**, 190–195 (1985).

[14] V. I. Bardyshev, A. M. Velikanov, and S. G. Gershman, ''Experimental studies of underwater noise in the ocean,'' Sov. Phys. Acoust. **16**, 512–513 (1971).
[15] L. N. Zakharov and V. A. Kirshov, ''Undersea noise in a coastal region,'' in *Eight All-union Acoustics Conference* (Moscow, 1973).
[16] I. A. Svendsen, P. A. Madsen, and J. B. Hansen, ''Wave characteristics in the surf zone,'' in *Proceedings of the 16th coastal engineering conference* (Academic Society of Civil Engineers, New York, 1978), pp. 520–539.
[17] I. A. Svendsen, ''Surf zone dynamics,'' *IUTAM symposium*, Sydney, Australia (Eds. M. L. Banner and R. H. J. Grimshaw, Springer-Verlag, New York, 1991), pp. 55–68.
[18] R. L. Wiegel, *Oceanographical Engineering* (Prentice Hall, Engle Wood Cliffs, NJ, 1964).
[19] C. J. Galvin, ''Breaker classification on three laboratory beaches,'' J. Geophys. Res. **73**, 3651–3659 (1968).
[20] O. B. Wilson, *et al.*, ''Noise source level density due to surf—part I: Monterey Bay, CA,'' IEEE Journal of Oceanic Engineering **22**, 425–433 (1997).
[21] S. A. Thorpe, ''On the clouds of bubbles formed by breaking wind-waves in deep water, and their role in air-sea gas transfer,'' Philos. Trans. R. Soc. London, Ser. A **304**, 155–210 (1982).
[22] S. A. Thorpe, ''A model of the turbulent diffusion of bubbles below the sea surface,'' J. Phys. Oceanogr. **14**, 841–854 (1984).
[23] W. M. Carey, J. W. Fitzgerald, and D. G. Browning, ''Low frequency noise from breaking waves,'' in *Natural Physical Sources of Underwater Sound* (Kluwer Academic, Dordrecht, 1993), pp. 277–304.
[24] E. C. Monahan and M. Lu, ''Acoustically relevant bubble assemblages and their dependence on meteorological parameters,'' IEEE J. Ocean Eng. **15**, 340–349 (1990).
[25] H. Medwin, ''*In situ* acoustic measurements of bubble populations in coastal ocean waters,'' J. Geophys. Res. **75**, 599–611 (1970).
[26] B. D. Johnson and R. C. Cooke, ''Bubble populations and spectra in coastal waters: A photographic approach,'' J. Geophys. Res. **84**, 3761–3766 (1979).
[27] H. Medwin and N. D. Breitz, ''Ambient and transient bubble spectral densities in quiescent seas and under spilling breakers,'' J. Geophys. Res. **94**, 12 751–12 759 (1989).
[28] S. Baldy, ''Bubbles in the close vicinity of breaking waves: Statistical characteristics of the generation and dispersion mechanism,'' J. Geophys. Res. **93**, 8239–8248 (1988).
[29] C. Lin and H. H. Hwung, ''External and internal flow fields of plunging breakers,'' Exp. Fluids **12**, 229–237 (1992).
[30] T. G. Leighton, A. D. Phelps, and D. G. Ramble, ''Acoustic bubble sizing: From the laboratory to surf zone trials,'' Acoust. Bull. 5–13 (1996).
[31] M. Strasberg, ''Nonacoustic noise interference in measurements of infrasonic ambient noise,'' J. Acoust. Soc. Am. **66**, 1487–1493 (1979).
[32] R. A. Saenger, ''Power spectrum of ''surfseisms,'' J. Acoust. Soc. Am. **32**, 1497 (1960).
[33] R. A. Saenger, ''''Surfseisms—Elastic waves generated in the bottom by pounding surf,'' J. Acoust. Soc. Am. **32**, 915 (1960).
[34] R. A. Saenger, ''An estimate of the offshore ambient noise spectrum produced by pounding surf,'' J. Acoust. Soc. Am. **33**, 1674 (1961).
[35] P. D. Thorne, ''Laboratory and marine measurements on the acoustic detection of sediment transport,'' J. Acoust. Soc. Am. **80**, 899–910 (1986).
[36] P. D. Thorne, ''Seabed saltation noise,'' in *Natural Physical Sources of Underwater Sound* (Kluwer Academic, Dordrecht, 1993), pp. 721–744.
[37] G. J. Franz, ''Splashes as sources of sound in liquids,'' J. Acoust. Soc. Am. **31**, 1080–1096 (1959).
[38] J. A. Nystuen, ''Rainfall measurements using underwater ambient noise,'' J. Acoust. Soc. Am. **74**, 972–982 (1986).
[39] H. C. Pumphrey and L. A. Crum, ''Acoustic emissions associated with drop impacts,'' in *Sea Surface Sound* (Kluwer Academic, Dordrecht, 1988), pp. 463–483.
[40] H. N. Oguz and A. Prosperetti, ''Bubble entrainment by the impact of drops on liquid surfaces,'' J. Fluid Mech. **219**, 143–179 (1990).
[41] H. N. Oguz and A. Prosperetti, ''Drop impact and the underwater noise of rain,'' in *Natural Physical Sources of Underwater Sound* (Kluwer Academic, Dordrecht, 1993), pp. 669–682.
[42] M. J. Lighthill, ''On sound generated aerodynamically II. Turbulence as a source of sound,'' Proc. R. Soc. London, Ser. A **222**, 1–32 (1954).
[43] A. P. Dowling and J. E. Ffowcs Williams, *Sound and Sources of Sound* (Ellis Horwood, Chichester, 1983).
[44] A. Prosperetti, ''Bubble dynamics in oceanic ambient noise,'' in *Sea Surface Sound* (Kluwer Academic, Dordrecht, 1987), pp. 151–171.
[45] H. L. Grant, V. M. Hughes, V. M. Vogel, and A. Moilliet, ''The spectrum of temperature fluctuations in turbulent flow,'' J. Fluid Mech. **34**, 423–442 (1968).
[46] T. Osborn, S. Vagle, D. Farmer, M. Cure, S. Thorpe, and A. Hall, ''Bubble plumes and turbulence,'' in *Natural Physical Sources of Underwater Sound* (Kluwer Academic, Dordrecht, UK, 1993), pp. 519–525.
[47] A. Prosperetti, ''Bubble-related ambient noise in the ocean,'' J. Acoust. Soc. Am. **84**, 1042–1054 (1988).
[48] W. M. Carey and D. Browning, ''Low-frequency ocean ambient noise: measurement and theory,'' in *Natural Mechanisms of Surface Generated Noise in the Ocean* (Kluwer Academic, Dordrecht, 1988), pp. 361–376.
[49] A. Kolaini, R. Roy, and L. A. Crum, ''An investigation of the acoustic emissions from a bubble plume,'' J. Acoust. Soc. Am. **89**, 2452–2455 (1991).
[50] A. R. Kolaini and L. A. Crum, ''Observations of underwater sound from laboratory breaking waves and the implications concerning ambient noise in the ocean,'' J. Acoust. Soc. Am. **96**, 1755–1765 (1994).
[51] A. R. Kolaini, R. A. Roy, and D. L. Gardner, ''Low-frequency acoustic emissions in fresh and salt water,'' J. Acoust. Soc. Am. **96**, 1766–1772 (1994).
[52] L. d'Agostino and E. Brennen, ''Acoustical absorption and scattering cross sections of spherical bubble clouds,'' J. Acoust. Soc. Am. **84**, 2126–2134 (1988).
[53] M. L. Banner and D. H. Cato, ''Physical mechanisms of noise generation by breaking waves—A laboratory study,'' in *Sea Surface Sound* (Kluwer Academic, Dordrecht, 1988), pp. 429–436.
[54] H. Medwin and M. M. Beaky, ''Bubble sources of the Knudsen sea noise spectra,'' J. Acoust. Soc. Am. **86**, 1124–1130 (1989).
[55] I. N. Didenkulov and A. S. Korotkov, ''The measurement of underwater noise characteristics in shallow water,'' in *Natural Physical Sources of Underwater Sound* (Kluwer Academic, Dordrecht, 1993), pp. 45–49.
[56] M. S. Longuet-Higgins, ''Bubble noise mechanisms—A review,'' in *Natural Physical Sources of Underwater Sound* (Kluwer Academic, Dordrecht, 1993), pp. 419–452.
[57] P. A. Hwang, Y. Poon, and J. Wu, ''Temperature effects on generation and entrainment of bubbles induced by a water jet,'' J. Phys. Oceanogr. **21**, 1602–1605 (1991).
[58] M. Koga, ''Bubble entrainment in breaking wind waves,'' Tellus **34**, 481–489 (1982).
[59] P. C. M. Jansen, ''Laboratory observations of the kinematics in the aerated region of breaking waves,'' Coastal Eng. **9**, 453–477 (1986).
[60] P. A. Crowther, ''Bubble noise creation mechanisms,'' in *Sea Surface Sound* (Kluwer Academic, Dordrecht, 1988), pp. 131–150.
[61] C. S. Clay and H. Medwin, *Acoustical Oceanography: Principles and Applications* (Wiley, New York, 1977).
[62] A. B. Wood, *A Textbook of Sound* (Bell, London, 1964), pp. 610.
[63] M. Minnaert, ''On musical air-bubbles and the sounds of running water,'' Philos. Mag. **16**, 235–248 (1933).
[64] K. W. Commander and A. Prosperetti, ''Linear pressure waves in bubbly liquids: Comparison between theory and experiments,'' J. Acoust. Soc. Am. **85**, 732–746 (1989).
[65] S. Finette and R. M. Heitmeyer, ''Angle-time-frequency resolution of the noise field generated by wind-induced breaking waves,'' J. Acoust. Soc. Am. **99**, 209–221 (1996).
[66] F. P. Shepard, *Submarine Geology* (Harper & Row, New York, 1973).
[67] M. J. Buckingham, ''Theory of three-dimensional acoustic propagation in a wedgelike ocean with a penetrable bottom,'' J. Acoust. Soc. Am. **82**, 198–210 (1987).
[68] S. A. L. Glegg and G. B. Deane, ''Experimental verification of the theory for sound propagation in a wedge with a shear supporting bottom,'' J. Acoust. Soc. Am. **92**, 2301–2302 (1992).
[69] G. B. Deane and M. J. Buckingham, ''An analysis of the three-dimensional sound field in a penetrable wedge with a stratified fluid or elastic basement,'' J. Acoust. Soc. Am. **93**, 1319–1328 (1993).
[70] E. K. Westwood, ''Broadband modeling of the three-dimensional penetrable wedge,'' J. Acoust. Soc. Am. **92**, 2212–2222 (1992).
[71] L. M. Brekhovskikh, ''Waves in layered media,'' in *Applied Mathematics and Mechanics* (Academic, San Diego, 1980).

The variation of modal wave numbers with geoacoustic parameters in layered media

Ronald T. Kessel
School of Earth and Ocean Sciences, Petch Building 191, University of Victoria, P.O. Box 3055, Victoria, British Columbia V8W 3P6, Canada

(Received 12 November 1996; revised 21 July 1997; accepted 25 July 1997)

Analytic expressions are derived for the variation of modal wave numbers due to variations in five basic properties of a layered environment: the compressional wave speed, shear wave speed, density, and thickness of any layer, and the depth of an interface between layers. Both fluid and solid layers are permitted. The layers may absorb energy, and the modal wave numbers and their variations are generally complex valued. The analytic expressions provide an efficient way to linearize the relation between the modal wave numbers and many geoacoustic parameters, for use in field sensitivity studies, range-dependent propagation models, partially linearized matched-field inversion schemes, and in the study of modes generally. © *1997 Acoustical Society of America.* [S0001-4966(97)02911-1]

PACS numbers: 43.30.Bp, 43.30.Ma, 43.30.Pc [SAC-B]

INTRODUCTION

In the analysis of physical systems it is in many ways useful to simplify matters by linearizing the relation between dependent and independent variables. In ocean acoustics, when the ocean can be treated as a layered elastic medium and the sound field as the superposition of its normal modes of vibration, linearization entails the partial derivatives of the modal wave numbers k with respect to any of five basic environmental properties: three elastic properties—the compressional (P) wave speed α, shear (S) wave speed β, and density ρ of a layer; and two geometric properties, one related to the other—the thickness h of a layer, and the depth z of an interface between layers.

Derivatives with respect to the environmental parameters turn up in many applications, in field sensitivity studies for range-independent media,[1] for example, and in range-dependent media, when modeling sound propagation through horizontally refracting ''mode rays.'' [2,3] They also turn up in matched-field inversion for range-dependent environments, although they may not be explicitly identified as such, when mode look-up tables are constructed to efficiently approximate the modes for a family of probable layer profiles using linear interpolation.[4] They are also central to a perturbative inversion method based on normal modes.[5]

The derivatives can be approximated numerically for many-layered media using a mode search program, together with a finite-difference scheme, by computing the perturbation in the wave number δk for a mode due to a small perturbation $\delta\epsilon_l$ made to some property, $\epsilon=\alpha$, β, ρ, h, or z, of the lth layer,

$$\frac{\partial k}{\partial \epsilon_l}=\lim_{\delta\epsilon\to 0}\frac{\delta k}{\delta\epsilon_l}\approx\frac{\delta k}{\delta\epsilon_l}. \tag{1}$$

Unfortunately, numerical differentiation can be imprecise, and it becomes time consuming when there are many modes and many environmental properties. Analytic expressions are preferred, being both more precise and faster, but the set of five expressions—$\partial k/\partial\alpha_l$, $\partial k/\partial\beta_l$, $\partial k/\partial\rho_l$, $\partial k/\partial h_l$, $\partial k/\partial z_l$—is apparently incomplete. Expressions for the elastic parameters α, β, and ρ are available from seismology,[6] but expressions for the geometric, layer thickness h, and interface depth z have not been reported when solid layers are involved.

Derivatives for all five geoacoustic parameters are given here for media having both fluid and solid layers. The media may absorb energy, and the modal wave numbers and their derivatives are in general complex valued. A uniform approach is followed that encompasses a number of important mode properties, including mode orthogonality, Rayleigh's principle, group velocity, and the excitation of a mode by a source.[7] Of these, only Rayleigh's principle and group velocity are derived here to illustrate the method used for the wave number derivatives. The method builds on the analysis of waves in layered media, of which there are many detailed treatments.[6–9] A brief outline following Kennett's versatile notation[10] is given to show how the elastic field in range-independent layered media can be represented as the superposition of cylindrical waves whose resonances are the normal modes of vibration. A key integral relation for modes is then formulated from which Rayleigh's principle, group velocity, and the five wave number derivatives are all derived.

An alternative approach, perhaps more elegant for lossless media, might begin with Hamilton's principle for elastic waves together with the calculus of variations.[11] But its extension to media that absorb energy, in which the field variables, elastic parameters, and modal wave numbers are generally complex, entails difficulties of its own. The advantage of the present method is that it is at once applicable to lossless and lossy media, and that it builds on concepts and equations that should be familiar to mode-propagation modelers who might be inclined to make use of the final expressions.

I. MODES IN LAYERED MEDIA

A. Fundamental equations

It is assumed that the medium is piecewise homogeneous, consisting of homogeneous fluid and solid layers stacked in any order, each layer being in contact with its neighbor along a plane horizontal interface. In an oceanic model, for example, the seawater is represented by a band of fluid layers whose elastic properties approximate the vertical variation of the water column in a stepped fashion. Below the seawater, there are solid layers representing stratified seafloor sediments, and above there is an upper interface with the atmosphere, possibly with solid layers interposed between the air and water representing surface ice.

In a cylindrical polar coordinate system (r,ϕ,z), with the z-axis pointing vertically downward, the displacement field $\mathbf{u}$ of the elastic waves can be written as

$$\mathbf{u}(r,\phi,z,t)=u_r(r,\phi,z,t)\hat{\mathbf{e}}_r+u_\phi(r,\phi,z,t)\hat{\mathbf{e}}_\phi + u_z(r,\phi,z,t)\hat{\mathbf{e}}_z, \tag{2}$$

where $\hat{\mathbf{e}}_r$, $\hat{\mathbf{e}}_\phi$, and $\hat{\mathbf{e}}_z$ are the unit vectors of the coordinate system. The equations of motion can be written as

$$\begin{aligned}
&\rho\partial_{tt}u_r=\partial_z\tau_{rz}+\partial_r\tau_{rr}+r^{-1}\partial_\phi\tau_{r\phi}+r^{-1}(\tau_{rr}-\tau_{\phi\phi})+\rho f_r,\\
&\rho\partial_{tt}u_\phi=\partial_z\tau_{\phi z}+\partial_r\tau_{r\phi}+r^{-1}\partial_\phi\tau_{\phi\phi}+2r^{-1}\tau_{r\phi}+\rho f_\phi,\\
&\rho\partial_{tt}u_z=\partial_z\tau_{zz}+\partial_r\tau_{rz}+r^{-1}\partial_\phi\tau_{\phi z}+r^{-1}\tau_{rz}+\rho f_z,
\end{aligned} \tag{3}$$

where τ_{ij} are elements of the stress tensor, ρ is the media density, and the first and second partial derivatives with respect to time t and position are denoted as ∂_i and ∂_{ij}, with i, $j=t$, r, ϕ, or z; and where $\mathbf{f}=[f_r,f_\phi,f_z]^T$ is the body force generating the waves, which may be set to zero in what follows because a mode, in any linear physical system, is by definition an oscillation that persists in the absence of forcing.

Hooke's law for locally homogeneous and isotropic media can be written as

$$\begin{aligned}
&\tau_{rr}=(\lambda+2\mu)\partial_r u_r+\lambda\{\partial_z u_z+r^{-1}(\partial_\phi u_\phi+u_r)\},\\
&\tau_{\phi\phi}=(\lambda+2\mu)r^{-1}(\partial_\phi u_\phi+u_r)+\lambda(\partial_z u_z+\partial_r u_r),\\
&\tau_{zz}=(\lambda+2\mu)\partial_z u_z+\lambda\{\partial_r u_r+r^{-1}(\partial_\phi u_\phi+u_r)\},\\
&\tau_{rz}=\mu(\partial_z u_r+\partial_r u_z),\\
&\tau_{\phi z}=\mu(r^{-1}\partial_\phi u_z+\partial_z u_\phi),\\
&\tau_{r\phi}=\mu(\partial_r u_\phi+r^{-1}\partial_\phi u_r-r^{-1}u_\phi),
\end{aligned} \tag{4}$$

where λ and μ are the Lamé coefficients.

B. Transformation to cylindrical waves

It can be shown that, within a layer, Eqs. (3) and (4) give the first-order differential equation[7,10,12]

$$\partial_z\begin{bmatrix}u_z\\ u_V\\ \tau_{zz}\\ \tau_{Vz}\end{bmatrix}=\begin{bmatrix}0 & -\lambda/\lambda_* & 1/\lambda_* & 0\\ -\Delta_h & 0 & 0 & 1/\mu\\ \rho\partial_{tt} & 0 & 0 & -1\\ 0 & \rho(\partial_{tt}-\nu^2\Delta_h) & -(\lambda/\lambda_*)\Delta_h & 0\end{bmatrix}\times\begin{bmatrix}u_z\\ u_V\\ \tau_{zz}\\ \tau_{Vz}\end{bmatrix}, \tag{5}$$

where

$$\begin{aligned}
&u_V=r^{-1}[\partial_r(ru_r)+\partial_\phi u_\phi],\\
&\tau_{Vz}=r^{-1}[\partial_r(r\tau_{rz})+\partial_\phi\tau_{\phi z}],\\
&\lambda_*=\lambda+2\mu,\\
&\nu^2=4\lambda(\lambda+\mu)/\rho\lambda_*,\\
&\Delta_h\equiv r^{-1}\partial_r(r\partial_r)+r^{-2}\partial_{\phi\phi},
\end{aligned} \tag{6}$$

have been introduced for convenience. The subscript V emphasizes that this is just one of two independent systems of equations, called P-SV waves, representing coupled P and S waves. The other equations, omitted here, represent pure shear wave motion in solid layers, called SH waves, whose vibrations do not involve fluid layers in the absence of layer roughness and scattering inclusions.

Differentiation with respect to t, r, and ϕ can be converted to multiplication in Eq. (5) by applying the Fourier–Bessel transform

$$\hat{\psi}(k,m,z,\omega)\equiv\frac{1}{2\pi}\int_{-\infty}^{+\infty}dt e^{+i\omega t}\int_0^{+\infty}dr r J_m(kr)\times\int_{-\pi}^{+\pi}d\phi e^{-im\phi}\psi(r,\phi,z,t), \tag{7}$$

to u_z, u_V, τ_{zz}, and τ_{Vz}, denoted generally here as ψ. This transform represents the elastic field as the superposition of cylindrical P-SV waves, each characterized by a frequency ω, horizontal wave number k, and azimuthal order m. Let us assume that the field is axially symmetric about the z axis. Then $m=0$ is the only azimuthal order of interest because $\hat{\psi}(k,m\neq 0,z,\omega)=0$. This does not limit the analysis because, in our ideal layered model, azimuthal variation can only enter through the source excitation $\mathbf{f}$, which is immaterial here.

In the transform domain, it is convenient to work interchangeably with the horizontal wave number k and slowness $p=k/\omega$; to use the compressional and shear phase speeds, $\alpha=\sqrt{(\lambda+2\mu)/\rho}$ and $\beta=\sqrt{\mu/\rho}$, rather than λ and μ; and to introduce the scaled variables

$$\begin{aligned}
&U=\hat{u}_z,\quad P=\omega^{-1}\hat{\tau}_{zz},\\
&V=-k^{-1}\hat{u}_V,\quad S=-(\omega k)^{-1}\hat{\tau}_{Vz}.
\end{aligned} \tag{8}$$

The transformation of Eq. (5) then becomes

$$\partial_z \begin{bmatrix} U \\ V \\ P \\ S \end{bmatrix}$$

$$= \omega \begin{bmatrix} 0 & p(1-2\beta^2/\alpha^2) & (\rho\alpha^2)^{-1} & 0 \\ -p & 0 & 0 & (\rho\beta^2)^{-1} \\ -\rho & 0 & 0 & p \\ 0 & \rho(vp^2-1) & -p(1-2\beta^2/\alpha^2) & 0 \end{bmatrix} \times \begin{bmatrix} U \\ V \\ P \\ S \end{bmatrix}, \tag{9}$$

with

$$v = 4\beta^2(1-\beta^2/\alpha^2). \tag{10}$$

This result must be modified for fluid layers where $\beta \to 0$. An ideal inviscid fluid cannot support shear stress, making $\tau_{rz} = \tau_{\phi z} = \tau_{Vz} = S = 0$ and reducing Eq. (9) to

$$\partial_z \begin{bmatrix} U \\ P \end{bmatrix} = \omega \begin{bmatrix} 0 & (\alpha^{-2} - p^2)/\rho \\ -\rho & 0 \end{bmatrix} \begin{bmatrix} U \\ P \end{bmatrix}, \tag{11}$$

with V dependent on P

$$\rho V = -pP. \tag{12}$$

The transform variables may be collected into a *stress-displacement* vector $\mathbf{b}(z) = [U,V,P,S]^T$, which in turn may be partitioned into to displacement $\mathbf{w}(z)$ and stress $\mathbf{t}(z)$ related vectors, $\mathbf{b}(z) = [\mathbf{w}(z), \mathbf{t}(z)]^T$, with

$$\mathbf{w}(z) = \begin{cases} [U,V]^T & \text{in solid layers,} \\ [U,V]^T \text{ or } [U] & \text{in fluid layers,} \end{cases}$$
$$\mathbf{t}(z) = \begin{cases} [P,S]^T & \text{in solid layers,} \\ [P,0]^T \text{ or } [P] & \text{in fluid layers.} \end{cases} \tag{13}$$

For convenience, the vector $\mathbf{b}$, its components U, V, P, and S, and its partitions $\mathbf{w}$ and $\mathbf{t}$, will all be used interchangeably in what follows. Both the two- or four-vector form of $\mathbf{b}$ will be required for fluid layers; the smaller being simply a subset of the larger.

C. An eigenvalue value problem

Whether in fluid or solid layers, the stress-displacement vector $\mathbf{b}(z)$ satisfies an equation of the form

$$[\partial_z - \omega A(p,z)]\mathbf{b}(z) = \mathbf{0}; \tag{14}$$

A being the matrix in Eq. (9) for solid layers, and Eq. (11) for fluid.

The elements of $\mathbf{b}(z)$ are coupled between layers by continuity conditions applied to the elastic field in the following way: U, like u_z, is continuous across all interfaces; V, like u_r and u_ϕ, is continuous across solid–solid interfaces in welded contact, but discontinuous at fluid interfaces owing to slip boundary conditions; and P and S, like τ_{zz}, τ_{rz}, and $\tau_{\phi z}$, are continuous across all interfaces, while S also vanishes at fluid–solid interfaces owing to slip boundary conditions.

If there are impenetrable upper or lower boundaries to the layered medium, then the elements of $\mathbf{b}(z)$ satisfy ''natural'' boundary conditions. Among the most common boundaries are an upper traction-free boundary with the atmosphere, where $P = S = 0$ because $\tau_{rz} = \tau_{\phi z} = \tau_{zz} = 0$; a rigid boundary in welded contact with a solid layer, where $U = V = 0$ because $\mathbf{u} = \mathbf{0}$; and a rigid boundary in slip contact with a fluid layer, where $U = 0$ because $u_z = 0$. On the other hand, if the layered medium is infinitely thick, then *physical (proper)* solutions satisfy the radiation condition: $\mathbf{b}(z) \to \mathbf{0}$ as $|z| \to \infty$. *Leaky (improper)* solutions, which are usually of lesser importance to long-range propagation, violate the radiation condition and are omitted here.

Equation (14), together with field continuity and boundary conditions, specifies an eigenvalue problem whose eigenvalues, ω and p, are the frequency and horizontal slowness of the normal modes, and whose eigenfunctions $\mathbf{b}(z)$ represent the characteristic vibrations of each mode as a function of depth. There are many numerical methods for computing the series of eigenvalues p, generally complex valued, at a given frequency ω.[7,12–15] They will not be reviewed here. Let us simply assume that the modal slownesses p—or equivalently, the modal wave numbers $k = \omega p$—and their eigenfunctions $\mathbf{b}(z)$ are available one way or another for any layer profile of interest.

The modes (ω,k) and their vertical dependence $\mathbf{b}(z)$ for cylindrical waves are the same as those for two-dimensional plane waves whose field variables are invariant in one horizontal direction. Equations (9)–(13) are in fact recovered in two dimensions (x,z) using the Fourier transform

$$\hat{\psi}(k,z,\omega) \equiv \frac{1}{2\pi} \int_{-\infty}^{+\infty} dt\ e^{+i\omega t} \int_{-\infty}^{+\infty} dx\ e^{-ikx} \psi(x,z,t), \tag{15}$$

in place of Eq. (7), and

$$U = i\hat{u}_z, \quad P = i\omega^{-1}\hat{\tau}_{zz},$$
$$V = \hat{u}_x, \quad S = \omega^{-1}\hat{\tau}_{xz}. \tag{16}$$

in place of Eq. (8). The present analysis therefore applies equally for both the three- and two-dimensional settings.

II. AN INTEGRAL IDENTITY FOR MODES

From Eq. (14) and the field continuity conditions, it follows that $\mathbf{b}(z)$ is a continuous function of depth in strictly solid media, and in strictly fluid media when its two-vector form is used. In combined solid and fluid media, however, $\mathbf{b}(z)$ is discontinuous across every fluid–solid interface, where the dimensions of $\mathbf{b}(z)$ jump from two to four when its two-vector form is used in fluids, or across every interface with a fluid layer, where V is discontinuous when its four-vector form is used in fluids.

Despite these isolated discontinuities, the inner product

$$\mathbf{w}(z) \cdot \mathbf{t}(z) = U(z)P(z) + V(z)S(z) \tag{17}$$

is always continuous because S is continuous everywhere and vanishes in a fluid, thereby eliminating discontinuous V. If $\mathbf{w}(z)$ and $\mathbf{t}(z)$ are the displacement and stress components of the same eigenfunction, then the inner product is closely related to the vertical energy flux in the mode.[10,7] But more important for the present analysis, Eq. (17) is also continuous if $\mathbf{w}(z)$ and $\mathbf{t}(z)$ are taken from the eigenfunctions of two different modes in the same layered environment, as if for the derivation of mode orthogonality, for example, using $\mathbf{w}_a(z)$ from mode a having $(\omega,k)=(\omega_a,k_a)$ for example, and $\mathbf{t}_b(z)$ from mode b having $(\omega,k)=(\omega_b,k_b)$. More generally still, Eq. (17) remains continuous if, as a purely mathematical exercise, the two modes a and b are taken from two *different* layered environments, a and b, provided that the fluid–solid interfaces of both environments are aligned in depth, and that z lies at a depth inside both.

Given *any* two modes a and b, then, we can construct an integral identity for the dot product, much as we could for any continuous function:

$$\omega_b[\mathbf{w}_a(z_2)\cdot\mathbf{t}_b(z_2+\Delta z)-\mathbf{w}_a(z_1)\cdot\mathbf{t}_b(z_1+\Delta z)] = \omega_b\int_{z_1}^{z_2}\frac{\partial}{\partial z}\,(\mathbf{w}_a(z)\cdot\mathbf{t}_b(z+\Delta z))dz. \tag{18}$$

Here the subscripts a and b identify properties associated with each mode, Δz is an arbitrary vertical shift of eigenfunction b relative to a (possibly to align a fluid–solid interface between the two environments), z_1 and z_2 lie within layered medium a supporting mode a, and $z_1+\Delta z$ and $z_2+\Delta z$ lie within a different layered medium b supporting mode b. This fundamental identity is the basis for the derivations that follow.

A. Rayleigh's principle

The simplest application of Eq. (18) occurs when the environment, frequency, and modal wave number are the same for modes a and b. Setting $\Delta z=0$, then expanding the integrand, eliminating all vertical derivatives using Eq. (9), and integrating from top $z_1=z_0$ to bottom $z_2=z_L$ of the media, where

$$\mathbf{w}_a(z_L)\cdot\mathbf{t}_b(z_L)=\mathbf{w}_a(z_0)\cdot\mathbf{t}_b(z_0)=0, \tag{19}$$

owing to the boundary or radiation conditions, we get the equivalent of Rayleigh's principle,

$$\omega^2 I_1=I_2, \tag{20}$$

in which

$$I_1=\int_{z_0}^{z_L}\rho(U^2+V^2)dz,$$
$$I_2=\int_{z_0}^{z_L}\left[\omega^2\left(\frac{P^2}{\rho\alpha^2}+\frac{S^2}{\rho\beta^2}\right)+\rho k^2 v V^2\right]dz. \tag{21}$$

That is, the kinetic energy $\omega^2 I_1$ in a mode equals the potential energy I_2.[6,10] Subscripts a and b have been dropped here because they denote the same mode. When integrating through fluid layers, in Eqs. (21) and elsewhere, substitute for V using Eq. (12) and set $S=v=0$.

B. Group velocity

The group velocity for a mode can be derived by perturbing the frequency of mode b from that of a, $\omega_a=\omega$, and $\omega_b=\omega+\delta\omega$, while keeping the same environment for both. The modal slowness will be perturbed along the dispersion curve for the mode, $p_a=p$ and $p_b=p+\delta p$, and likewise the eigenfunctions, $\mathbf{b}_a=\mathbf{b}$ and $\mathbf{b}_b=\mathbf{b}+\delta\mathbf{b}$. With these substitutions, Eq. (18) can be applied both as it is written, and with subscripts a and b interchanged, giving two different identities. Expanding the integrands of both, eliminating all vertical derivatives using Eq. (9), then taking the difference of both identities retaining terms to first order, and finally integrating from top z_0 to bottom z_L, we find

$$0=I_1\delta\omega-pI_3(p\,\delta\omega+\omega\,\delta p)=I_1\delta\omega-pI_3\delta k, \tag{22}$$

where

$$I_3=\int_{z_0}^{z_L}\left[\frac{1}{p}\,US-\frac{1}{p}\left(1-2\frac{\beta^2}{\alpha^2}\right)VP+\rho v V^2\right]dz,$$
$$\delta k=p\,\delta\omega+\omega\,\delta p. \tag{23}$$

The group velocity is therefore

$$c_g=\frac{\partial\omega}{\partial k}=\lim_{\delta k\to 0}\frac{\delta\omega}{\delta k}=\frac{pI_3}{I_1}. \tag{24}$$

III. MODE PERTURBATIONS DUE TO α, β, AND ρ

The mode sensitivities to α, β, and ρ can now be derived in a similar way. Again, let mode b be the perturbed mode a, but now the perturbation is due to small changes in α, β, and ρ, making environment b slightly different from a, although the depths of the layer interfaces are, for the moment, the same for both. Let the elastic parameters for mode a be $\epsilon_a(z)=\epsilon(z)$, and those for b be $\epsilon_b(z)=\epsilon(z)+\delta\epsilon(z)$, with $|\delta\epsilon(z)|\ll|\epsilon(z)|$ and $\epsilon=\alpha$, β, or ρ. The slowness of mode b will be perturbed from that of a, $p_a=p$ and $p_b=p+\delta p$, but the frequency will remain unchanged, $\omega_a=\omega_b=\omega$. Once again construct two forms of Eq. (18)—first as written, then with subscripts a and b interchanged. Expanding their integrands and eliminating all vertical derivatives using Eq. (9), then taking the difference of both identities retaining terms to first order, and finally solving for δk we find

$$\delta k=\omega\,\delta p=\int_{z_0}^{z_L}[s_\alpha(z)\,\delta\alpha+s_\beta(z)\,\delta\beta+s_\rho(z)\,\delta\rho]dz, \tag{25}$$

in which

$$s_\alpha(z)=\frac{\omega}{pI_3}\,[4\beta^2 p(PV-\rho p\beta^2 V^2)-P^2/\rho]/\alpha^3,$$
$$s_\beta(z)=-\frac{\omega}{pI_3}\,[4\beta\rho p^2(1-2\beta^2/\alpha^2)V^2+4\beta pPV/\alpha^2 + S^2/(\beta^3\rho)], \tag{26}$$
$$s_\rho(z)=\frac{1}{2}\,\frac{\omega}{pI_3}\,[U^2+V^2-P^2/(\alpha\rho)^2-S^2/(\beta\rho)^2 - p^2 v V^2],$$

are *sensitivity densities*: a depth-dependent weighting that the vibrations of a mode (U, V, P, S) confer to environmental perturbations apportioning their influence on the mode's wave number. In fluid layers, set $s_\beta = 0$, $S = v = 0$, then $\beta = 0$.

The first-order approximation to the wave number perturbation Eq. (25) is, of course, the linear combination of perturbations due to each elastic parameter considered independently. It depends, moreover, as the variation of any acoustic-field variable should as a rule depend, on both the size of the perturbation and on the span of depths where the perturbations are made relative to the energy distribution in the field, for δk is insensitive to perturbations made at depths where $s_\epsilon(z)$ is relatively weak, but sensitive where $s_\epsilon(z)$ is relatively strong, which corresponds roughly to the depths where the strength of the mode (U, V, P, S) is accordingly weak or strong. A plot of $s_\epsilon(z)$ as a function of depth therefore indicates the depths where the mode is most sensitive to variations in $\epsilon = \alpha$, β, or ρ.

Notice that $s_\rho(z)$ can be written as

$$s_\rho(z) = \frac{1}{2\omega p I_3 \rho(z)} (\omega^2\, dI_1 - dI_2), \tag{27}$$

in which $dI_{1,2}$ are the integrands of $I_{1,2}$ in Eq. (21). Scaling the density uniformly in all layers [$\delta\rho(z)/\rho(z) = \overline{\delta\rho} = \text{constant}$], while holding the phase speeds α and β constant ($\delta\alpha = \delta\beta = 0$), does not perturb the wave number because

$$\delta k = \int_{z_0}^{z_L} s_\rho(z)\, \delta\rho\, dz = \frac{1}{2\omega p I_3} (\omega^2 I_1 - I_2)\, \overline{\delta\rho} = 0, \tag{28}$$

due to Rayleigh's principle (20).

Also notice that there are no restrictions on the thickness of the layers, so Eqs. (25) and (26) apply in the limit as the thickness of the layers is made vanishingly small, when the medium varies continuously with depth.

The partial derivative of the wave number with respect to just one elastic parameter in a single homogeneous layer l follows from Eq. (25) by making perturbations to each parameter in turn, applied uniformly throughout the layer, from just below its upper interface $z = z_{l-1}^+$ to just above its lower interface z_l^-,

$$\frac{\partial k}{\partial \epsilon_l} = \lim_{\delta\epsilon \to 0} \frac{\delta k}{\delta \epsilon} = \int_{z_{l-1}^+}^{z_l^-} s_\epsilon(z)\, dz; \quad \epsilon = \alpha, \beta, \rho. \tag{29}$$

The integral in Eq. (29) can only be evaluated after ω,k, and $\mathbf{b}(z)$ have been determined for the mode, which must be done numerically for all but the simplest layered media as mentioned earlier. But the integration through a single layer can in any case be carried out analytically for greater computational speed and precision. Upon further analysis using matrix methods, it can be shown that $\mathbf{b}(z)$ is a linear combination of up and down traveling P and S waves,[7,8,10] whose vertical dependencies go as complex exponential functions $e^{\pm i\gamma z}$, typical of plane or cylindrical waves (γ being the vertical wave number for P or S waves, which are independent of depth within a layer). The vertical dependence of $s_\epsilon(z)$ is therefore exponential as well, and it may be integrated straightforwardly (although the algebra proves tedious), giving $\partial k/\partial \epsilon_l$ directly in terms of $\mathbf{b}(z_{l-1}^+)$ and $\mathbf{b}(z_l^-)$.

A. Perturbations due to energy absorption

Perturbations made to the rate of energy absorption by the media are implicit in the perturbations made to α and β because increasing the absorption rate amounts to adding a small imaginary part to the phase speeds.[9] Assuming, for example, an harmonic time dependence $e^{-i\omega t}$ for the waves and an energy absorption rate of η nepers per wavelength for the homogeneous media within a layer, the complex phase speed for either wave type is given by $c = c_r - i c_i$, where c_r is the real part of the phase speed, and $c_i \approx \eta c_r/(2\pi)$ is the imaginary.[7] Perturbations in absorption $\delta\eta$ are therefore reflected in phase speed as $\delta c = -i\,\delta c_i \approx -i\,\delta\eta c_r/(2\pi)$.

IV. MODE PERTURBATIONS DUE TO LAYER THICKNESS h_l AND INTERFACE DEPTH z_l

Now let the perturbation to the layered media be a slight increase δh in the thickness of the l'th homogeneous layer, while keeping all other properties unchanged. $|\delta h|$ must be small compared with the natural length scales of variation in the perturbed layer of which there are two: (1) the vertical variation in the eigenfunction $\mathbf{b}(z)$, being of the order of the shear wave length $\lambda_\beta = 2\pi\beta/\omega$ in a solid layer, and compressional wavelength $\lambda_\alpha = 2\pi\alpha/\omega$ in a fluid; and (2) the vertical variation of the media, being conservatively estimated as the thickness h_l of the layer. (The latter may be somewhat larger when many thin layers are used to approximate gradual continuous variation in a stepwise manner.)

If the upper boundary z_0 of the media is held fixed, then the perturbation pushes down all interfaces and layers below the upper interface of layer l by an amount δh. The perturbed medium is therefore

$$\begin{aligned} &\epsilon_b(z + H(z - z_l^-)\,\delta h) = \epsilon_a(z), \quad \text{for } z_0 < z < z_L, \\ &\epsilon_b(z) = \epsilon_a(z_l^-), \quad \text{for } z_l^{-1} \leqslant z < z_l^- + \delta h, \end{aligned} \tag{30}$$

for $\epsilon = \alpha$, β, and ρ. Here $H(z - z_l^-)$ is the Heaviside step function, and z_l^- is the depth of the lower boundary of layer l, taken just above the boundary, inside that layer. The perturbation changes the modal slowness, $p_a = p$ and $p_b = p + \delta p$, but the frequency remains unchanged $\omega_a = \omega_b = \omega$.

The perturbation in the eigenfunction is

$$\delta\mathbf{b}(z) = \mathbf{b}_b(z + H(z - z_l^-)\,\delta h) - \mathbf{b}_a(z), \quad \text{for } z_0 < z < z_L. \tag{31}$$

Although $\mathbf{b}(z)$ and $\delta\mathbf{b}(z)$ may be discontinuous functions of z at a fluid interface, $\delta\mathbf{b}(z)$ is nevertheless a continuous function of δh, as required by the perturbation method, because the layer interfaces below the perturbed layer are, so far as the $\mathbf{b}_a$ and $\mathbf{b}_b$ are concerned, kept in alignment by the stepped shift in the argument of $\mathbf{b}_b$.

Substituting mode a and the perturbed mode b into the integral identity (18) with $\Delta z = H(z - z_l^-)\,\delta h$, then integrating in two steps through the entire thickness of media a, taking care to split the integration at $z = z_l^-$, where $\mathbf{t}_b(z + H(z - z_l^-)\,\delta h)$ is discontinuous owing to the step function, we find

$$-\omega[\mathbf{w}_a(z_l^-)\cdot(\mathbf{t}_b(z_l^-+\delta h)-\mathbf{t}_b(z_l^-))]$$
$$=\omega\int_{z_0}^{z_l^-}\frac{\partial}{\partial z}[\mathbf{w}_a(z)\cdot\mathbf{t}_b(z)]dz$$
$$+\omega\int_{z_l^-}^{z_L}\frac{\partial}{\partial z}[\mathbf{w}_a(z)\cdot\mathbf{t}_b(z+\delta h)]dz. \tag{32}$$

The left side has been simplified using

$$\mathbf{w}_a(z_0)\cdot\mathbf{t}_b(z_0)=\mathbf{w}_a(z_L)\cdot\mathbf{t}_b(z_L+\delta h)=0, \tag{33}$$

because both eigenfunctions satisfy the same boundary or radiation conditions. A first-order Taylor series approximation can be used for eigenfunction $\mathbf{b}_b$ because $|\delta h|$ is small,

$$\mathbf{b}_b(z_l^-+\delta h)\approx\mathbf{b}_b(z_l^-)+\left(\frac{\partial}{\partial z}\mathbf{b}_b(z_l^-)\right)\delta h, \tag{34}$$

in which case Eq. (32) becomes

$$-\omega\mathbf{w}_a(z_l^-)\cdot\frac{\partial}{\partial z}\mathbf{t}_b(z_l^-)\,\delta h$$
$$=\omega\int_{z_0}^{z_L}\frac{\partial}{\partial z}[\mathbf{w}_a(z)\cdot\mathbf{t}_b(z+H(z-z_l^-)\,\delta h)]dz. \tag{35}$$

Now, repeating the derivation of Eq. (35) with the roles of mode a and b reversed in Eq. (18), we find a similar relation:

$$-\omega\mathbf{t}_a(z_l^-)\cdot\frac{\partial}{\partial z}\mathbf{w}_b(z_l^-)\,\delta h$$
$$=\omega\int_{z_0}^{z_L}\frac{\partial}{\partial z}[\mathbf{w}_b(z+H(z-z_l^-)\,\delta h)\cdot\mathbf{t}_b(z)]dz. \tag{36}$$

Taking the difference of Eqs. (35) and (36), substituting $\mathbf{b}_a(z)=\mathbf{b}(z)$ and $\mathbf{b}_b=\mathbf{b}(z)+\delta\mathbf{b}(z)$ from Eq. (31), then expanding the integrands, eliminating all vertical derivatives using Eq. (9), and retaining terms to first order, then finally solving for δk, we find

$$\delta k=s_h(z_l^-)\,\delta h, \tag{37}$$

where

$$s_h(z)=\frac{\omega}{2pI_3}\{p[2(1-2\beta^2/\alpha^2)PV-2US-vp\rho V^2]$$
$$+\rho(U^2+V^2)+[(P/\alpha)^2+(S/\beta)^2]/\rho\} \tag{38}$$

is the partial derivative with respect to layer thickness itself because

$$\frac{\partial k}{\partial h_l}=\lim_{\delta h\to 0}\frac{\delta k}{\delta h}=s_h(z_l^-). \tag{39}$$

The interior depth z_l^- at which a homogeneous layer is widened should be inconsequential. Indeed, it can be shown that

$$\frac{\partial s_h(z)}{\partial z}=0 \tag{40}$$

inside a layer, hence z_l^- may be taken anywhere inside the lth layer to give the same result. It follows that, if $\mathbf{b}(z)$ vanishes at any depth inside the layer, then so do $s_h(k,z)$ and the corresponding δk for that layer, which is to say that a mode is insensitive to perturbations to the thickness of a layer in which the eigenfunction vanishes. This happens when the mode's vibrations decay exponentially before traversing the entire layer, as *physical* (*proper*) modes invariably do in infinitely thick layers, for example. Insensitivity under these conditions makes sense because we can always imagine making the layer thicker at depths undetected by the mode's vibrations.

A somewhat different perturbation, in which the depth z_l of the common interface between layer l and $l+1$ is perturbed without changing the depths of any other interfaces, can be derived using Eq. (37). In effect, two opposing perturbations of layer thickness are made simultaneously in adjacent layers—the upper layer l becoming thicker by an amount δh, and the lower $l+1$ becoming thinner by the same amount. The net perturbation is simply the sum of the two thickness variations considered separately

$$\delta k=[s_h(z_l^-)-s_h(z_{l+1}^-)]\,\delta h. \tag{41}$$

The partial derivative with respect to the interface depth z_l is therefore

$$\frac{\partial k}{\partial z_l}=\frac{\partial k}{\partial h_{l-1}}-\frac{\partial k}{\partial h_l}=s_h(z_l^-)-s_h(z_{l+1}^-). \tag{42}$$

V. CONCLUSION

The partial derivatives of the modal wave number k with respect to any geoacoustic parameter, α_l, β_l, ρ_l, h_l, or z_l, can be computed efficiently for any homogeneous layer l using analytic expression (26) in (29), and (38) in (39) and (42). Having thus obtained the partial derivatives, the effect of any finite perturbation to one parameter in the l'th layer can be approximated as

$$\delta k\approx(\partial k/\partial\epsilon_l)\,\delta\epsilon_l;\quad \epsilon=\alpha,\ \beta,\ \rho,\ h,\ \text{and}\ z, \tag{43}$$

and the effect of many simultaneous perturbations as the sum of each acting alone. When tracking the wave numbers for local modes in range-dependent environments, for example, in which the layer profile is a function of both range and depth, $\epsilon_l=\epsilon_l(r,z)$, the net perturbation due to a small range step δr is

$$\delta k=\left(\frac{\partial k}{\partial r}\right)\delta r=\sum_{\alpha_l,\beta_l,\rho_l,h_l}\frac{\partial k}{\partial\epsilon_l}\frac{\partial\epsilon_l(r,z)}{\partial r}\,\delta r, \tag{44}$$

where the summation is over all four parameters in all layers.

The effect of many perturbations is perhaps best handled using a *mode sensitivity matrix* $\mathcal{S}$, whose elements are the partial derivatives of k_n for each mode n, to each parameter ϵ_l in each layer,

$$\mathcal{S}_{nl}=\partial k_n/\partial\epsilon_l. \tag{45}$$

Then the vector $\delta\mathbf{k}$ holding the perturbation to all modes due to any perturbation of the layered media, when also represented as a vector $\delta\epsilon$ whose elements are $\delta\epsilon_l$ for $\epsilon=\alpha$, β, ρ, and h, in layers $l=1,...,L$, is

$$\delta\mathbf{k}\approx\mathcal{S}\delta\epsilon. \tag{46}$$

Note that $\mathscr{S}$ is generally not square. It will have as many rows as there are modes, and as many columns as there are perturbed environmental parameters. Also note that $\delta\mathbf{k}$ may include modes at any number of frequencies, giving any number of rows.

Equation (46) finds application in matched-field inversion, when the properties of the environment are inferred from the sound field. For by inverting $\mathscr{S}$, using a least squares method if need be, it is possible to compute an environmental ''correction'' $\delta\epsilon$ to be applied to a first-guess layer profile in order to shift its modes a specified amount $\delta\mathbf{k}$, calculated to improve the match between measured and modeled data. In principle, this is an extension to all geoacoustic parameters of the perturbative inversion method that Rajan *et al.*[5] developed to invert for α and ρ as continuous functions of depth in seafloor sediments in the absence of shear waves, but Eq. (46) assumes discrete layering rather than continuous variation of the sediments.

[1] J. M. Ozard and M. L. Yeremy, ''Effects of mismatch on matched-field processing for Arctic and Pacific shallow water,'' 22nd mtg. TTCP GTP9 Subgroup G, 18–22 October 1993.

[2] B. E. McDonald, ''Bathymetric and volumetric contributions to ocean acoustic mode coupling,'' J. Acoust. Soc. Am. **100**, 219–224 (1996).

[3] R. Burridge and H. Weingberg, ''Horizontal rays and vertical modes,'' in *Lecture Notes in Physics*, Vol. 70, Wave Propagation and Underwater Acoustics, edited by J. B. Keller and J. S. Papadakis (Springer-Verlag, Berlin, 1977).

[4] S. E. Dosso, ''Efficient acoustic field computations for matched field inversion for geoacoustic propagation,'' J. Acoust. Soc. Am. **97**, 3293–3293 (1995).

[5] S. D. Rajan, J. F. Lynch, and G. V. Frisk, ''Perturbative inversion methods for obtaining bottom geoacoustic parameters in shallow water,'' J. Acoust. Soc. Am. **82**, 998–1017 (1987).

[6] H. Takeuchi and M. Saito, ''Seismic surface waves,'' in *Methods in Computational Physics*, Vol. 11, edited by B. A. Bolt (Academic, New York, 1972).

[7] R. T. Kessel, ''Scattering of elastic waves in layered media: a boundary integral–normal mode method,'' Ph.D. dissertation, University of Victoria, Victoria, BC (1996).

[8] L. Brekhovskihk, *Waves in Layered Media* (Academic, New York, 1980), 2nd ed.

[9] K. Aki and P. R. Richards, *Quantitative Seismology* (Freeman, San Francisco, 1980), Vol. 1.

[10] B. L. Kennett, *Seismic Wave Propagation in Stratified Media* (Cambridge U.P., Cambridge, 1983).

[11] J. H. Woodhouse, ''On Rayleigh's principle,'' Geophys. J. R. Astron. Soc. **46**, 11–22 (1976).

[12] T. W. Dawson and R. T. Kessel, ''SAMPLE: A seismo-acoustic mode program for layered environments. Theoretical Background,'' Defence Research Establishment Atlantic (Canada), Technical Memorandum DREA/TM/95/225 (Oct. 1995).

[13] M. A. Porter and E. L. Reiss, ''A numerical method ocean-acoustic normal modes,'' J. Acoust. Soc. Am. **76**, 244–251 (1984).

[14] J. I. Arvelo, M. Talmant, and H. Uberall, ''A normal mode model of underwater propagation including shear effects in a layered ocean floor,'' Comput. Acoust. **3**, 131–150 (1990).

[15] C. T. Tindle and N. R. Chapman, ''A phase function for finding normal mode eigenvalues over a layered elastic bottom,'' J. Acoust. Soc. Am. **96**, 1777–1782 (1994).

Wave-number estimation in an ocean waveguide

Leonid G. Krasny and Sergey P. Antonyuk
Institute of Hydromechanics of Ukrainian Academy of Sciences, Zhelyabov Str., 8/4 Kiev, 252057, Ukraine

(Received 14 August 1995; revised 4 April 1997; accepted 3 July 1997)

The optimal (by the criterion of maximum likelihood function) algorithm is applied to the problem of estimating the wave numbers of normal modes in an ocean waveguide. This algorithm has a high-resolution property, takes into account availability of correlated spatial noise, does not require prior modal decorrelation, is suitable for a horizontal array of any shape, and permits one to attain the potential wave-number estimation accuracy limit (Cramer–Rao bound). Numerical simulations for the Pekeris model of an ocean waveguide are presented, demonstrating that the optimal algorithm can essentially (more than ten times) improve the wave-number estimation performance relative to those one for the MUSIC algorithm. Wave-number estimation by the optimal algorithm requires searching for the global extremum of some goal function. It is connected, in common, with known computing difficulties and the absence of sufficiently well-developed computing algorithms. Algorithms based on the signal eigenvectors property of the spectral correlation matrix measured on the array elements are free from listed drawbacks. The synthesis of such algorithms and an asymptotic analysis of their performance are accomplished under conditions of arbitrary correlated spatial noise. Among these algorithms there is one that permits the estimation of the wave numbers as accurately as the optimal algorithm. Experimental verification of this result has been received for the most cases of practical interest.
[S0001-4966(97)04210-0]

PACS numbers: 43.30.Bp, 43.30.Wi, 43.30.Ma [SAC-B]

INTRODUCTION

The wave-number estimation of normal modes propagating in an ocean waveguide is a problem of considerable interest in ocean acoustics. Having the wave-number estimates, it becomes possible to solve a set of important inverse hydroacoustic problems, such as ocean acoustic tomography,[1,2] source localization,[3] and waveguide characterization.[4]

For normal mode wave-number estimation, various algorithms of spatial processing of the field received by a hydrophone array have been proposed. The matched-field processing algorithm was the first to emerge. This algorithm assumes usage of the array as a beamformer in the wave-number space. The most important property of this algorithm is the simplicity of its technical realization, with the most essential drawback being a low-resolution threshold, which is inversely proportional to the array's aperture. If the resolution requirement is high but the array length is limited, then a certain kind of algorithm possessing high-resolution ability can be tried for this purpose.[5–7] So, the Prony method has been suggested for the modal wave-number estimation problem.[4] But a practical application of this method is limited by uniformly spaced line arrays and high signal-to-noise ratio. Later, Candy and Sullivan also used the high-resolution algorithm MUSIC for this purpose.[8] This algorithm entails eigendecomposition of the array signal correlation matrix and uses the property of its noise subspace vectors. The MUSIC algorithm is applicable to arbitrary array geometry and relatively low signal-to-noise ratio. Simultaneously, accuracy of the wave-number estimates received by this method essentially depends on the intermodal correlation coefficient.[9–11] In the case of a deterministic waveguide, the modes are coherent[11] (fully correlated) and the MUSIC algorithm is not applicable.[11,12] Chouhan and Anand[11] have proposed decorrelating the modes by horizontally moving the array in the waveguide. However, it is not made clear how the array moving length, i.e., the length on which the array is moved, affects the wave-number estimation accuracy and not how this accuracy can be improved, if it is impossible to move the array.

The Cramer–Rao bound (CRB) is known to characterize a potential wave-number estimation accuracy. While performing under some regular conditions, the CRB can be attained by the optimal (in the sense of maximum likelihood ratio) algorithm. In this paper we obtain this optimal algorithm, analyze its accuracy characteristics, and compare them with those one for the most known algorithms. All results are valid for the case of arbitrary correlated spatial noise. The purpose of this paper is to demonstrate the application of the optimal algorithm to the wave-number estimation problem.

A brief outline of this paper is now presented. In Sec. I the wave-number estimation problem is formulated. Assuming that the field received by the array is Gaussian, the optimal wave-number estimation algorithm is synthesized in Sec. II. Using the standard statistical theory of a maximum likelihood estimator, in Sec. III we derive an explicit asymptotic formula for CRB and discuss the opportunity of attaining this bound by the optimal algorithm. Section IV provides a formula demonstrating a clear relation between the array moving length and the statistical errors of the MUSIC algorithm. Simulation results given in Sec. V and related to the optimal and the MUSIC algorithms illustrate that the proposed optimal algorithm can essentially improve the wave-number estimation performance relative to those

one for the MUSIC algorithm. As shown in Sec. VI, so-called[10,13,14] weighted subspace fitting (WSF) algorithm based, such as the MUSIC algorithm, on the use of eigenvectors' properties can compete with the optimal algorithm if accumulation time, i.e., the time interval during which the field is received by the array, is large enough. Finally, Sec. VII presents some numerical results, illustrating the effectiveness of both the optimal and the WSF algorithms under conditions where the accumulation time is limited.

I. STATEMENT OF THE PROBLEM

Consider a plane-parallel waveguide in which a point source and a horizontal receiving array with N elements (hydrophones) are placed. Assume that a random narrow-band source is located in the Fraunhofer zone (far field) of the array and in the same plane as the array. The Fourier transform of the field $u(t,\mathbf{r}_n)$ received by the nth array element at the time interval $[t+(l-1),\ t+lT]$ has the form[15,16]

$$U_l(\omega;\mathbf{r}_n)=a_l(\omega)\sum_{m=1}^{M}A_{lm}(\omega)(k_m|\rho_n-\rho_0|)^{-1/2}\times\exp(jk_m|\rho_n-\rho_0|)+N_l(\omega;\mathbf{r}_n), \tag{1}$$

where $a_l(\omega)$ is the source spectral amplitude, $A_{lm}(\omega)\equiv A_{lm}(\omega;\mathbf{r}_0,z_a)$ are the modal amplitudes, $\mathbf{r}_0=(\rho_0,z_0)$ and $\mathbf{r}_n=(\rho_n,z_a)$ are, respectively, coordinates of the source and the array elements, k_m is the mth modal wave number, M is the number of propagating modes, and $N_l(\omega;\mathbf{r}_n)$ is the noise field spectral amplitude at the nth hydrophone, assumed to be uncorrelated with the source signal. It is convenient to rewrite relation (1) in the following matrix notation:

$$\mathbf{U}_l=a_l\mathbf{S}\mathbf{A}_l+\mathbf{N}_l, \tag{2}$$

where $\mathbf{U}_l=[U_l(\omega;\rho_1),\ldots,U_l(\omega;\rho_N)]^{\mathsf{T}}$ and $\mathbf{N}_l=[N_l(\omega;\rho_1),\ldots,N_l(\omega;\rho_N)]^{\mathsf{T}}$ are $N\times 1$ vectors, $a_l\equiv a_l(\omega)$, $\mathbf{S}=[\mathbf{S}(k_1),\ldots,\mathbf{S}(k_M)]$ is a $\mathbf{N}\times\mathbf{M}$ matrix whose mth column $\mathbf{S}(k_m)$ is the array response on the unit amplitude signal propagating in the mth mode, the elements $S_n(k_m)$ of the vector $\mathbf{S}(k_m)$ are equal to $S_n(k_m)=(k_m|\rho_n-\rho_0|)^{-1/2}\exp(jk_m|\rho_n-\rho_0|)$; $\mathbf{A}_l=[A_{l1}(\omega),\ldots,A_{lM}(\omega)]$ is a $M\times 1$ vector of the mode amplitudes.

The array signal correlation matrix is then given by

$$\mathbf{R}_{\mathrm{SN}}=\langle\mathbf{U}_l\mathbf{U}_l^{+}\rangle=\mathbf{S}\mathbf{P}\mathbf{S}^{+}+\mathbf{R}_N, \tag{3}$$

where $\langle\cdot\rangle$ denotes the expectation operator, the superscript $+$ denotes the Hermitian transpose, $\mathbf{P}=g_S\mathbf{A}_l\mathbf{A}_l^{+}$, $g_S\equiv g_S(\omega)=\langle a_l(\omega)a_l(\omega)^{*}\rangle$ is the source spectral power density, $*$ denotes the conjugate operator and $\mathbf{R}_N=\langle\mathbf{N}_l\mathbf{N}_l^{+}\rangle$ is the spatial correlation matrix of the waveguide and array noises.

The problem is to estimate the wave numbers k_m, having the observation vectors $\mathbf{U}_l$ for the L time intervals $[t+(l-1)T,\ t+lT]$.

A popular way of solving this problem is to use the matched field processing algorithm, which has the following form:

$$F_{\mathrm{conv}}(k)=\frac{\mathbf{S}^{+}(k)\mathbf{R}_N^{-1}\hat{\mathbf{R}}\mathbf{R}_N^{-1}\mathbf{S}(k)}{\mathbf{S}^{+}(k)\mathbf{R}_N^{-1}\mathbf{S}(k)}, \tag{4}$$

where $\mathbf{S}(k)$ is a $N\times 1$ vector with elements $S_n(k)=(k|\rho_n-\rho_0|)^{-1/2}\exp(jk|\rho_n-\rho_0|)$; $\mathbf{R}_N^{-1}$ denotes $N\times N$ matrix, which is the inverse of the $\mathbf{R}_N$ matrix; and

$$\hat{\mathbf{R}}=\frac{1}{L}\sum_{l=1}^{L}U_lU_l^{+} \tag{5}$$

is an estimate of the spatial correlation matrix $\mathbf{R}_{\mathrm{SN}}$ of the field at the aperture of the array.

In accordance with algorithm (4), the wave numbers can be estimated by locating the peaks of $F_{\mathrm{conv}}(k)$. However, if the distance $\Delta k_m=|k_m-k_{m-1}|$ between neighboring wave numbers k_m and k_{m-1} decreases, these two peaks merge into one maxima and the wave numbers are not resolved. In this case, it becomes necessary either to use longer arrays to improve the resolution or to abandon the matched field processing algorithm and employ the so-called high-resolution algorithms.[5,7,17] The latter approach is used to solve the problem of wave-number estimation.

II. SYNTHESIS OF AN OPTIMAL ALGORITHM

From the standard statistical theory of estimation a method of solving the above formulated problem is found. It is a synthesis of an optimal (by the criterion of the maximum of likelihood function) wave-number estimation algorithm. We use this approach to synthesize such an optimal algorithm.

We begin by assuming that the observation vectors $\mathbf{U}_l$ are statistically independent at different time intervals and being normally distributed within the parameters $\langle\mathbf{U}_l\rangle=0$, $\langle\mathbf{U}_l\mathbf{U}_m^{+}\rangle=\mathbf{R}_{\mathrm{SN}}\delta_{lm}$, where δ_{lm} is the Kroneker delta.

Then, with the accurate up to inessential constants, the log-likelihood function (LF)[10,18] is

$$\ln\lambda(\mathbf{U}|\mathbf{k},\mathbf{P})=-\mathrm{Tr}(\ln\mathbf{R}_{\mathrm{SN}}+\mathbf{R}_{\mathrm{SN}}^{-1}\hat{\mathbf{R}}), \tag{6}$$

where $\mathbf{U}=[\mathbf{U}_1^{\mathsf{T}},\ldots,\mathbf{U}_L^{\mathsf{T}}]^{\mathsf{T}}$ is the $NL\times 1$ vector formed by the observation $\mathbf{U}_l$ vectors, $\mathbf{k}=[k_1,\ldots,k_M]^{\mathsf{T}}$ is a vector of estimated parameters, T is the symbol of transpose, and $\mathrm{Tr}(\cdot)$ is a trace operator.

Interestingly included for us in the wave number k_m, the logarithm LF (6), in general, depends on matrix $\mathbf{P}$. This matrix is unknown, usually because the modal spectral power density radiated by the source and the modal amplitudes are unknown.

Following the estimation strategy of maximum likelihood method (MLM), first we find the maximum likelihood (ML) estimate $\hat{\mathbf{P}}$ of the matrix $\mathbf{P}$. Then we transpose it into (6) to obtain the optimal (according to the criterion of the maximum of likelihood function) wave-number estimation algorithm.

Set the derivative of function (6) with respect to Pequal to zero and solve the appropriate equation. As a result we obtain (see for details Refs. 10 and 18)

$$\hat{\mathbf{P}}=\mathbf{Q}^{-1}\mathbf{S}^{+}\mathbf{R}_N^{-1}\hat{\mathbf{R}}\mathbf{R}_N^{-1}\mathbf{S}\mathbf{Q}^{-1}-\mathbf{Q}^{-1}, \tag{7}$$

where $\mathbf{Q}\equiv\mathbf{S}^{+}\mathbf{R}_N^{-1}\mathbf{S}$.

Note that the nonsingularity of matrix $\mathbf{Q}$ is necessary for the existence of the MI estimate (7). When the matrix $\mathbf{R}_N$ is nonsingular, this usually takes place; matrix $\mathbf{Q}$ is also nons-

ingular, if the rank of matrix $\mathbf{S}$ is equal to $\mathbf{M}$ [$\mathrm{rank}(\mathbf{S})=\mathbf{M}$]. This is possible if at first number M of propagating in the waveguide modes does not exceed the number N of array elements ($M \leqslant N$) and; second, the array geometry is chosen so that the linear independence of the vectors $\mathbf{S}(k_m)$ is ensured for arbitrary values of wave number k_m belonging to the interval $[k_{\min} k_{\max}]$ of their possible meanings.

Use the obtained estimate (7) and substitute it into logarithm LF (6). Making several transformations,[10,18] we have

$$\ln \lambda(\mathbf{U}|k) \equiv \ln \lambda(\mathbf{U}|k,\hat{\mathbf{P}}) = \mathrm{Tr}\{\mathbf{\Pi}_S \hat{\mathbf{R}} \mathbf{R}_N^{-1} - \ln(\mathbf{\Pi}_S \hat{\mathbf{R}} \mathbf{R}_N^{-1})\}, \quad (8)$$

where $\mathbf{\Pi}_S = \mathbf{S}\mathbf{Q}^{-1}\mathbf{S}^{+}\mathbf{R}_N^{-1}$ is a $N \times N$ projection matrix onto the subspace spanned by vectors $\mathbf{S}(k_m)$ associated with the wave number k_m.

It should be noted here that, when the noise is spatially uncorrelated (i.e., $\mathbf{R}_N = g_N \mathbf{I}$, where g_N is the spectral power density of the noise and $\mathbf{I}$ is a $N \times N$ identity matrix), formulas (7) and (8) coincide with those received by Ottersten[10] and Jaffer.[18]

In accordance with algorithm (8), the optimal (by the criterion of the maximum of likelihood function) wave-number estimation algorithm assumes estimation of the spatial correlation matrix $\mathbf{R}_{\mathrm{SN}}$ (5) of the field at the array aperture (using the observation vectors $\mathbf{U}_l$), calculation of the projection matrix $\mathbf{\Pi}_S$, formation of goal function (8) (taking into account noise spatial correlation matrix $\mathbf{R}_N$), and its global maximizing. The coordinates $\hat{\mathbf{k}} = [\hat{k}_1, \ldots, \hat{k}_M]^{\mathsf{T}}$ of this maximum give the required ML estimates of the normal modes wave numbers.

III. POTENTIAL ACCURACY OF THE OPTIMAL ESTIMATES

From the general theory of estimation, it is known that an important characteristic of the ML estimates $\hat{k}$ of the wave number k is the correlation matrix

$$\mathbf{K}_{\mathrm{ML}} = \langle (\hat{\mathbf{k}} - \mathbf{k})(\hat{\mathbf{k}} - \mathbf{k})^{\mathsf{T}} \rangle. \quad (9)$$

Under some regularity conditions, the ML estimates are known to be asymptotically ($L \to \infty$) efficient, or in other words, their correlation matrix of the errors (9) attains the limit correlation matrix $\mathbf{K}_{\mathrm{RK}}$ called the Cramer–Rao bound.

The essential regularity condition is the consistency of the ML estimates $\hat{P}$ and $\hat{k}$. This condition is fulfilled in the case considered by us. As a result, the formula

$$L\mathbf{K}_{\mathrm{RK}} = \lim_{L \to \infty} (L\mathbf{K}_{\mathrm{ML}}), \quad (10)$$

being very useful for calculating the CRB, is valid.[9]

The matrix $\mathbf{K}_{\mathrm{RK}}$ can be derived using the approach developed by Stoica[9] and Ottersten.[10] As a result, we have

$$\mathbf{K}_{\mathrm{RK}} = \frac{1}{2L} [\mathrm{Re}\{\mathbf{H} \odot (\mathbf{P}\mathbf{S}^{+}\mathbf{R}_{\mathrm{SN}}^{-1}\mathbf{S}\mathbf{P})^{\mathsf{T}}\}]^{-1}, \quad (11)$$

where

$$\mathbf{H} \equiv \mathbf{D}^{+}\mathbf{R}_N^{-1}\mathbf{\Pi}_0\mathbf{D}; \quad (12)$$

$\mathbf{D} = [d_1, \ldots, d_M]$ is a $N \times M$ matrix whose mth column d_m is the $N \times 1$ vector with elements $d_{nm} = \partial S_n(k)/\partial k$ calculated at the point $k = k_m$; $\mathbf{\Pi}_0 \equiv \mathbf{I} - \mathbf{\Pi}_S$; $\odot$ denotes element-by-element multiplication of two matrices.

Formula (11) generalizes the results obtained in Ref. 10 in the case of arbitrary correlated spatial noise and permits us to calculate the limiting wave-number estimation accuracy, which can be attained asymptotically by the optimal algorithm (8). In other words, from (10) and (11) it follows that

$$\mathbf{K}_{\mathrm{ML}} = \mathbf{K}_{\mathrm{RK}} + o(L^{-1}), \quad (13)$$

where $o(L^{-1})$ denotes a matrix which elements have the order higher than L^{-1}.

Relation (11) facilitates the comparison of the estimation accuracy of the optimal algorithm with those yielded by others algorithms, specifically by the MUSIC algorithm.

IV. MUSIC ALGORITHM

We now direct our attention to the MUSIC algorithm, which is one of the most popular high-resolution algorithms, by elucidating how much it concedes in the wave-number estimation accuracy to the optimal algorithm (8).

The MUSIC algorithm is based on the property of noise eigenvectors of the equation

$$\mathbf{R}_{\mathrm{SN}} \varphi_n = \lambda_n \mathbf{R}_N \varphi_n, \quad (14)$$

where $\lambda_1, \ldots, \lambda_N$ are eigenvalues arranged in the monotonically nonincreasing order ($\lambda_n \geqslant \lambda_{n+1} > 0$); $\varphi_1, \ldots, \varphi_N$ are associated eigenvectors.

It is known[12,17] that if the number, N, of the array elements are greater than the number M of propagating modes, then, by virtue of the matrix $\mathbf{R}_{\mathrm{SN}}$ structure (3), eigenvectors $\varphi_1, \ldots, \varphi_N$, can be separated into two groups: the signal eigenvectors $\varphi_1, \ldots, \varphi_{\tilde{M}}$ corresponding to the largest eigenvalues, and noise eigenvectors $\varphi_{\tilde{M}+1}, \ldots, \varphi_N$, corresponding to the smallest eigenvalues $\lambda_{\tilde{M}+1} = \cdots = \lambda_N = 1$ (moreover, $\lambda_{\tilde{M}} > \lambda_{\tilde{M}+1}$). The bound $\tilde{M}$ of this separation is determined by the rank of the matrix $\mathbf{P}$, i.e., $\tilde{M} = \mathrm{rank}(\mathbf{P})$.

If among the modes none is coherent and $\mathrm{rank}(\mathbf{P}) = M = \tilde{M}$, then the orthogonality condition

$$\mathbf{S}^{+}(k_m)\varphi_n = 0, \quad m = 1,\ldots,M; \quad n = M+1,\ldots,N \quad (15)$$

fulfilled.[7,12] This condition means that any noise eigenvector φ_n is orthogonal to any column $\mathbf{S}(k_m)$ of matrix $\mathbf{S}$.

Having used condition (15), the MUSIC algorithm was synthesized in Refs. 7 and 8. By this algorithm, the wave-number estimates $\hat{k}_m$ can be obtained as the coordinates of the M highest peaks of the function

$$F_M(k) = [\mathbf{S}^{+}(k)\hat{\mathbf{\Psi}}_N \hat{\mathbf{\Psi}}_N^{+}\mathbf{S}(k)]^{-1}, \quad (16)$$

where $\hat{\mathbf{\Psi}}_N^{+} = [\varphi_{M+1}, \cdots, \hat{\varphi}_N]$ is the $N \times (N-M)$ matrix made up from estimates $\hat{\varphi}_n$ of the noise eigenvalues φ_n. The estimates $\hat{\varphi}_n$ can be derived from the equation

$$\hat{\mathbf{R}}\hat{\varphi}_n = \hat{\lambda}_n \mathbf{R}_N \hat{\varphi}_n, \quad (17)$$

in which $\hat{\lambda}_1 \geqslant \hat{\lambda}_2 \geqslant \cdots \geqslant \hat{\lambda}_N$ are the noise eigenvalue estimates arranged in decreasing order.

It can be shown[19] that the estimates of $\hat{k}_m$ obtained by the MUSIC algorithm (16) are asymptotically unbiased (i.e., $\lim_{L\to\infty}\hat{k}_m = k_m$) and have the following correlation matrix:

$$\mathbf{K}_M = \langle (\hat{k}-k)(\hat{k}-k)^{\mathsf{T}} \rangle$$

$$= \frac{1}{2}\frac{1}{2L}\,[\mathrm{Re}\{\mathbf{H}\odot\mathbf{I}\}]^{-1}\,\mathrm{Re}\{\mathbf{H}\odot(\mathbf{P}^{-1}+\mathbf{P}^{-1}\mathbf{Q}\mathbf{P}^{-1})^{\mathsf{T}}\}$$

$$\times[\mathrm{Re}\{\mathbf{H}\odot\mathbf{I}\}]^{-1} + o(L^{-1}). \quad (18)$$

This closed-form expression generalizes the result obtained by Stoica and Nehorai[19] on the case of arbitrary correlated spatial noise. When the noise is spatially uncorrelated, i.e., $R_N = g_N\mathbf{I}$, formula (18) coincides with those received in Ref. 19.

We now assume that the source is placed in a deterministic waveguide. Then, as has been shown in Ref. 11, the orthogonality condition (15) is not fulfilled [$\mathrm{rank}(\mathbf{P}) = 1 < M$] and the MUSIC algorithm becomes inapplicable for the wave-number estimation. To change this situation and apply the MUSIC algorithm, it has been proposed[11] to decorrelate the modes at the expense of moving of the array. It can be shown that, if during moving, the array remains far enough from the source, then matrix $\mathbf{P}$ is transformed into the smoothed correlation matrix

$$\mathbf{P}_C = \mathbf{P}\odot\mathbf{W}_C, \quad (19)$$

where $\mathbf{W}_C$ is the $M\times M$ smoothing matrix with elements

$$[\mathbf{W}_C]_{nm} = \frac{\sin(L_b(k_n-k_m)/2)}{L_b(k_n-k_m)/2}\exp(jL_b(k_n-k_m)/2); \quad (20)$$

L_b denotes a moving interval.

The degree of modal decorrelation obviously depends on the value L_b of the moving interval. Nevertheless, the rank of matrix $\mathbf{P}$ remains equal to M, condition (15) is fulfilled, and for the wave-number estimation we can use the MUSIC algorithm.

The derived formulas (11) and (18) are now used to numerically investigate the estimation accuracy of the optimal algorithm (8) and the MUSIC algorithm (16).

V. NUMERICAL EXAMPLE

In this section, the performance of the algorithms described above is studied through computer simulation.

Let a horizontal equidistant linear array of $N=6$ hydrophones and a source radiating a narrow-band random Gaussian signal at a central frequency $f_0 = 40$ Hz be placed in a plane-parallel Pekeris waveguide (Fig. 1). Numerical values assigned to the parameters of the array and the waveguide are taken from Ref. 11 and are as follows: channel depth $h = 50$ m, source depth $z_0 = 25$ m, array depth $z_a = 25$ m, uniform interhydrophone spacing $|\rho_{n-1}-\rho_n| = 12.5$ m, sound speed in water $c_W = 1500$ m/s, sound speed in sediment $c_S = 2000$ m/s, density of the water layer $\rho_W = 1.0$ g/cm^3, density of the sediment half-space $\rho_S = 1.1$ g/cm^3. For the assigned values there are two propagating modes ($M=2$), whose wave numbers are equal to $k_1 = 0.1592$ and $k_2 = 0.1331$. The distance $|\rho_1-\rho_0|$ between the first (the nearest to the source) hydrophone and the source is equal to 1000 m. The noise is assumed to be delta correlated in space (spatially uncorrelated), so that the spatial correlation matrix is $\mathbf{R}_N = g_N\mathbf{I}$.

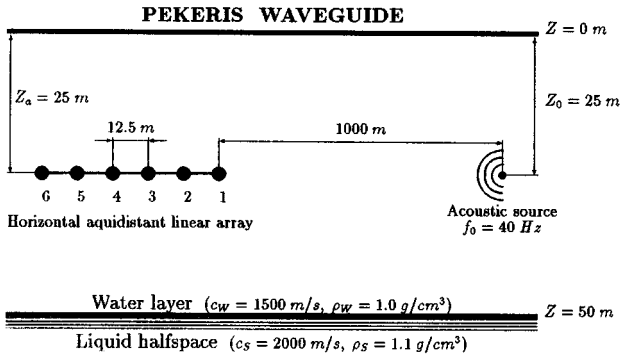


FIG. 1. Schematic placement of the array and source in the Pekeris waveguide. A six-element horizontal array with uniform interhydrophone spacing 12.5 m is positioned in the middle of waveguide. The source, positioned at the depth 25 m deep and distance 1000 m from the array hydrophone, radiates acoustic energy in a narrow frequency band about $f_0 = 40$ Hz.

Using Eqs. (11) and (18), the dependence of the root-mean-squared (rms) error σ_m of the wave-number estimates from the array moving interval L_b is calculated. The results of these calculations corresponding to the first wave number are $k_1 = 0.1592$, shown in Fig. 2. Curve 1 is obtained for the

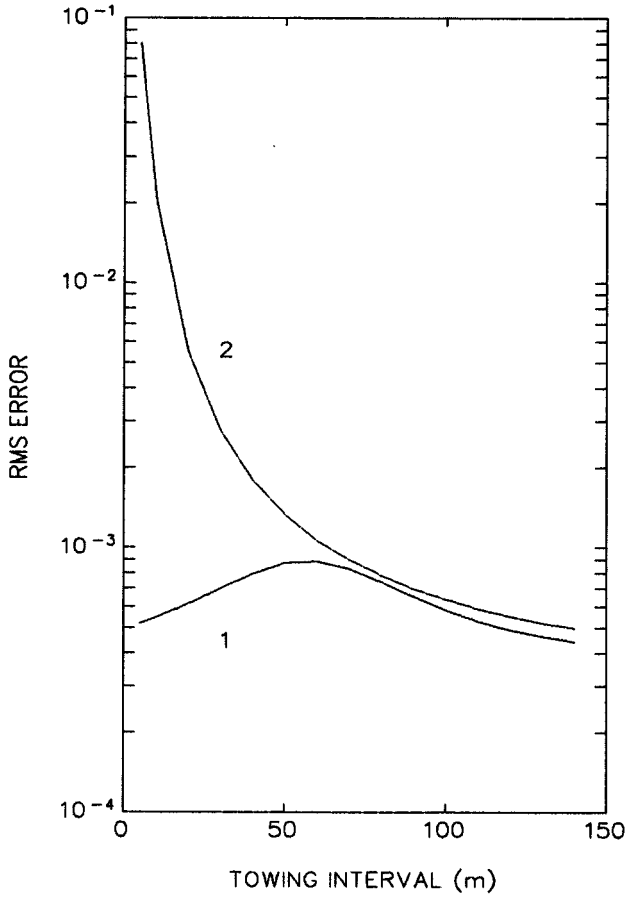


FIG. 2. rms error of the first wave number ($k_1 = 0.1592$) as a function of the moving interval. Curve 1 corresponds to the optimal algorithm and curve 2 to the MUSIC algorithm. SNR=60 dB, $L = 100$.

optimal algorithm (8) ($\sigma_1=[\mathbf{K}_{\mathrm{RK}}]_{11}^{1/2}$), curve 2 is obtained for the MUSIC algorithm (16) ($\sigma_1=[\mathbf{K}_M]_{11}^{1/2}$). Performing calculations based on Eq. (18), the smoothed correlation matrix $\mathbf{P}_C$ is used instead of matrix $\mathbf{P}$. The signal-to-noise ratio (SNR) g_S/g_N is equal to 60 dB. The average value of the SNR on the hydrophones $\tilde{g}_m=g_S\mathbf{S}^+(k_m)\mathbf{S}(k_m)p_{mm}/Ng_N$ ($m=1,2$) is equal to 15.5 dB for the first mode and 5.5 dB for the second one. The number of observation vectors $\mathbf{U}_l$ utilized for estimating $\hat{\mathbf{R}}$ is equal to 100, i.e., $L=100$.

Analysis of the curves displayed in Fig. 2 shows that the optimal algorithm gives more accurate wave-number estimates than the MUSIC algorithm. The distinction in rms error between these two algorithms increases as the moving interval and, as a consequence, the intermodal correlation coefficient decreases.

VI. WAVE-NUMBER ESTIMATION ALGORITHMS BASED ON THE SIGNAL EIGENVECTORS PROPERTY

As was demonstrated above, the MUSIC algorithm's wave-number estimation accuracy essentially depends on the value of the intermodal correlation coefficient. It is not always possible to decorrelate modes by moving the array in the medium, because in practice the restriction on the length of the moving interval L_b is often present. The optimal algorithm is free from this limitation, enabling evaluation of the wave numbers at arbitrary modal correlation, or, therefore, it does not requiring the array moving. However, for finding the wave-number estimates, the global extremum of the goal function (8) must be found, which is complicated with the known computing difficulties[20] and lack of sufficiently well-developed computing algorithms, which take into account the structure of the goal function. The algorithms based on the signal eigenvectors property of Eq. (14) are free from the listed drawbacks.

According to Ref. 12, when premultiplied by $\mathbf{R}_N$, each signal eigenvector $\varphi_1,\cdots,\varphi_{\tilde{M}}$ of Eq. (14) is a linear combination of vectors $\mathbf{S}(k_m)$, $m=1,...,M$, that is:

$$\mathbf{R}_N\varphi_n=\mathbf{S}(k_1)T_{1n}+\mathbf{S}(k_2)T_{2n}+\cdots+\mathbf{S}(k_M)T_{Mn},\quad n=1,...,\tilde{M}, \tag{21}$$

where T_{nm} are the elements of a $M\times\tilde{M}$ matrix $\mathbf{T}[\mathrm{rank}(\mathbf{T})=\tilde{M}\leqslant M]$. Expression (21) can be considered a system of algebraic equations. The number of these systems is equal to $\tilde{M}$, and every system is linear relative to the coefficients $T_{1n},\ldots,T_{Mn}$ and nonlinear relative to the wave numbers $k_1,\ldots,k_M$.

Because further calculations are more convenient when made in matrix form, (21) is presented as follows:

$$\mathbf{R}_N\mathbf{\Psi}_S=\mathbf{ST}, \tag{22}$$

where $\mathbf{\Psi}_S=[\varphi_1,\ldots,\varphi_{\tilde{M}}]$ is a $N\times\tilde{M}$ matrix.

In actual practice, instead of matrix $\mathbf{\Psi}_S$ we have its estimate $\hat{\mathbf{\Psi}}_S=[\varphi_1,\ldots,\varphi_{\tilde{M}}]$ composed by the signal eigenvector estimates $\hat{\varphi}_n$ derived from Eq. (17). Thus condition (22) is true only approximately, i.e., $\mathbf{R}_N\hat{\mathbf{\Psi}}_S\approx\mathbf{ST}$. Therefore, a natural estimation criterion is to find matrix $\mathbf{T}$ and vector $\mathbf{k}$ by minimizing the residual function

$$\begin{aligned}F_W(\mathbf{k},\mathbf{T})&=\|(\mathbf{R}_N\hat{\mathbf{\Psi}}_S-\mathbf{ST})\mathbf{W}^{1/2}\|_{\mathbf{R}_N^{-1}}\\&=\mathrm{Tr}([(\mathbf{R}_N\hat{\mathbf{\Psi}}_S-\mathbf{ST})\mathbf{W}^{1/2}]^+\\&\quad\times\mathbf{R}_N^{-1}[(\mathbf{R}_N\hat{\mathbf{\Psi}}_S-\mathbf{ST})\mathbf{W}^{1/2}]),\end{aligned} \tag{23}$$

where $\mathbf{W}^{1/2}$ is a square root of a nonsingular positively defined Hermitian matrix $\mathbf{W}$ which must be chosen.

It is not difficult to make sure that matrix $\hat{\mathbf{T}}=\mathbf{Q}^{-1}\mathbf{S}^+\hat{\mathbf{\Psi}}_S$ minimizes the residual function (23) as vector $\mathbf{k}$ is fixed. Substituting $\hat{\mathbf{T}}$ back into (23), after some transformations we get

$$F_W(\mathbf{k})=\mathrm{Tr}(\mathbf{\Pi}_0\mathbf{R}_N\hat{\mathbf{\Psi}}_S\mathbf{W}\hat{\mathbf{\Psi}}_S^+), \tag{24}$$

where $F_W(\mathbf{k})\equiv F_W(\mathbf{k},\hat{\mathbf{T}})$. In the particular case of a spatially uncorrelated noise, i.e., $\mathbf{R}_N=g_N\mathbf{I}$, formula (24) coincides with those received in Refs. 10, 13, and 14.

Thus we have established that for receiving estimates $\hat{k}_m$ of the wave numbers k_m, the global maximum of goal function (24) must be found. Moreover, unlike the MUSIC algorithm (16), the degree of intermodal correlation does not affect condition (21) which has been used as the basis for algorithm (24).

Algorithm (24) actually generates a whole class of algorithms related to the different weighting matrices $\mathbf{W}$. It can be shown by analogy with Ottersten[10] and Viberg[13,14] that the algorithms of this class give asymptotically unbiased wave-number estimates $\hat{k}_m$ (i.e., $\lim_{L\to\infty}\hat{k}_m=k_m$) and their correlation matrix $\mathbf{K}_W=\langle(\hat{\mathbf{k}}-\mathbf{k})(\hat{\mathbf{k}}-\mathbf{k})^T\rangle$ of errors, essentially depends on the choice of weighting matrix $\mathbf{W}$.

The most frequently occurring matrices are the following three weighting matrices:[10,13,14]

$$\mathbf{W}_1=[\mathbf{\Lambda}_S-\mathbf{I}]^2\mathbf{\Lambda}_S^{-1},\quad \mathbf{W}_2=\mathbf{\Lambda}_S-\mathbf{I},\quad \mathbf{W}_3=\mathbf{I}, \tag{25}$$

where $\mathbf{\Lambda}_S=\mathrm{diag}(\hat{\lambda}_1,\ldots,\hat{\lambda}_{\tilde{M}})$ is a diagonal matrix.

Transposing (25) into (24) and assuming the noise to be spatially uncorrelated, i.e., $\mathbf{R}_N=g_N\mathbf{I}$, we receive the WSF algorithm[10,13,14] when $\mathbf{W}=\mathbf{W}_1$, the deterministic ML algorithm[10,13] when $\mathbf{W}=\mathbf{W}_2$, and the MD-MUSIC algorithm[10,13] when $\mathbf{W}=\mathbf{W}_3$.

Using a calculation method developed in Refs. 10 and 13, it can be shown that the following inequality is correct:

$$\mathbf{K}_{\mathbf{W}}\geqslant\mathbf{K}_{\mathrm{WSF}}=\frac{1}{2L}\,[\mathrm{Re}\{\mathbf{H}\odot(\mathbf{PS}^+\mathbf{R}_{\mathrm{SN}}^{-1}\mathbf{SP})^{\mathrm{T}}\}]^{-1}+o(L^{-1}), \tag{26}$$

which means that the wave-number estimation accuracy is bounded below by matrix $\mathbf{K}_{\mathrm{WSF}}$. The equality sign in (26) takes place in two cases: (1) if instead of the weighting matrix $\mathbf{W}$ the weighting matrix $\mathbf{W}_1$ is taken corresponding to the WSF algorithm;[10,13] (2) if the source is placed in the deterministic waveguide with, as in consequence, the coherent modes, i.e., $\mathrm{rank}(\mathbf{P})=1$. Moreover, it is easy to verify that in the second case the weighting matrix $\mathbf{W}$ does not affect the wave-number estimation accuracy.

It should be noted here that Ottersten[10] and Viberg[13,14] have reported Eq. (26) for a special case $\mathbf{R}_N=g_N\mathbf{I}$.

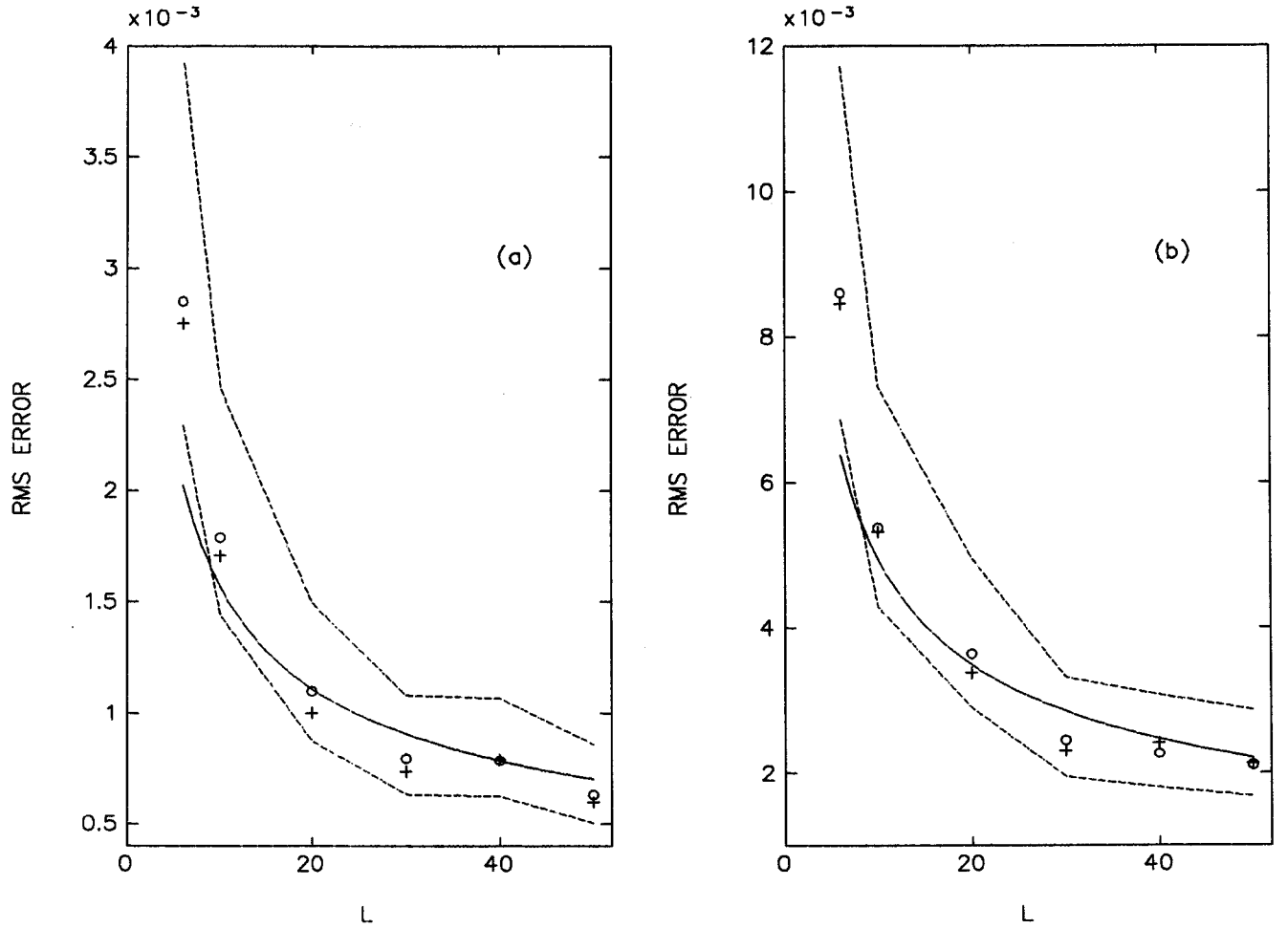


FIG. 3. rms errors of the first (a) and the second wave number (b) as functions of the number L of observation vectors. (+) the optimal algorithm; (○) the WSF algorithm; (—) Cramer–Rao bound; (---) 99% confidence interval for the optimal algorithm. SNR=60 dB, $L_b=0$ m.

Comparison of formulas (11), (13), and (26) shows that the WSF algorithm, which has been synthesized by taking into account the signal eigenvectors property (21), enables the attainment asymptotically ($L\to\infty$) as accurate wave-number estimates as those obtained by the optimal algorithm (8). Moreover, the accuracy of these estimates is a limited attainable accuracy, i.e., it is coinciding with the CRB matrix $\mathbf{K}_{\mathrm{RK}}$.

How the optimal algorithm and the WSF algorithm behave themselves when the number L of observation vectors is small is shown by making use of simulation experiments.

VII. SIMULATION EXPERIMENTS

Some results of the computer simulation of the optimal algorithm (8) and the WSF algorithm (24) are presented in this section. We are mainly interested in investigating the performance (in terms of bias and rms errors) of these two algorithms under conditions of limited accumulation time $T_o=LT$.

The calculations were accomplished for the previous example involving the plane-parallel waveguide and equidistance linear array. As demonstrated earlier, the signal field on the array was formed by two modes with wave numbers $k_1=0.1592$ and $k_2=0.1331$. The correlation matrix $\hat{\mathbf{R}}$ is evaluated according to formula (5), using L independent observation vector $\mathbf{U}_l=\mathbf{X}_h\boldsymbol{\xi}_l$, where $\mathbf{X}_h$ is the lower triangular matrix of the Cholesky decomposition of the matrix $\mathbf{R}_{\mathrm{SN}}$ ($\mathbf{R}_{\mathrm{SN}}=\mathbf{X}_h\mathbf{X}_h^{+}$); $\boldsymbol{\xi}_l$ are independent complex-valued Gaussian vectors with parameters $\langle\boldsymbol{\xi}_l\rangle=0$, $\langle\boldsymbol{\xi}_l\boldsymbol{\xi}_m^{+}\rangle=\mathbf{I}\delta_{lm}$. The noise is assumed to be delta correlated in space, i.e., $\mathbf{R}_N=g_N\mathbf{I}$. The SNR is equal to 60 dB. To provide a statistical basis for our conclusions, 50 independent trial runs were made for data samples $L=6,10,20,30,40,50$ ($L\geqslant N$). The sampled means and sampled rms errors are then calculated for every L.

To find the global minimum $\hat{\mathbf{k}}$ of the goal function (24), we have used the two-stage procedure described in Ref. 14. At stage one of this procedure, the initial estimates vector $\mathbf{k}^{(o)}$ is found by applying the coordinate descent method suggested in Refs. 17 and 21. At stage two, the Kaufman's modification of the Gauss–Newton method[14] were used to define the wave-number estimates more exactly. A similar two-stage procedure for the optimal algorithm (8) is developed.

Results of the computing experiments are shown in Fig. 3 and Fig. 4. The crosses mark the results relating to optimal algorithm (8) and the noughts to the WSF algorithm (24). The dashed lines demonstrate the 0.99 confidence intervals for the optimal algorithm.

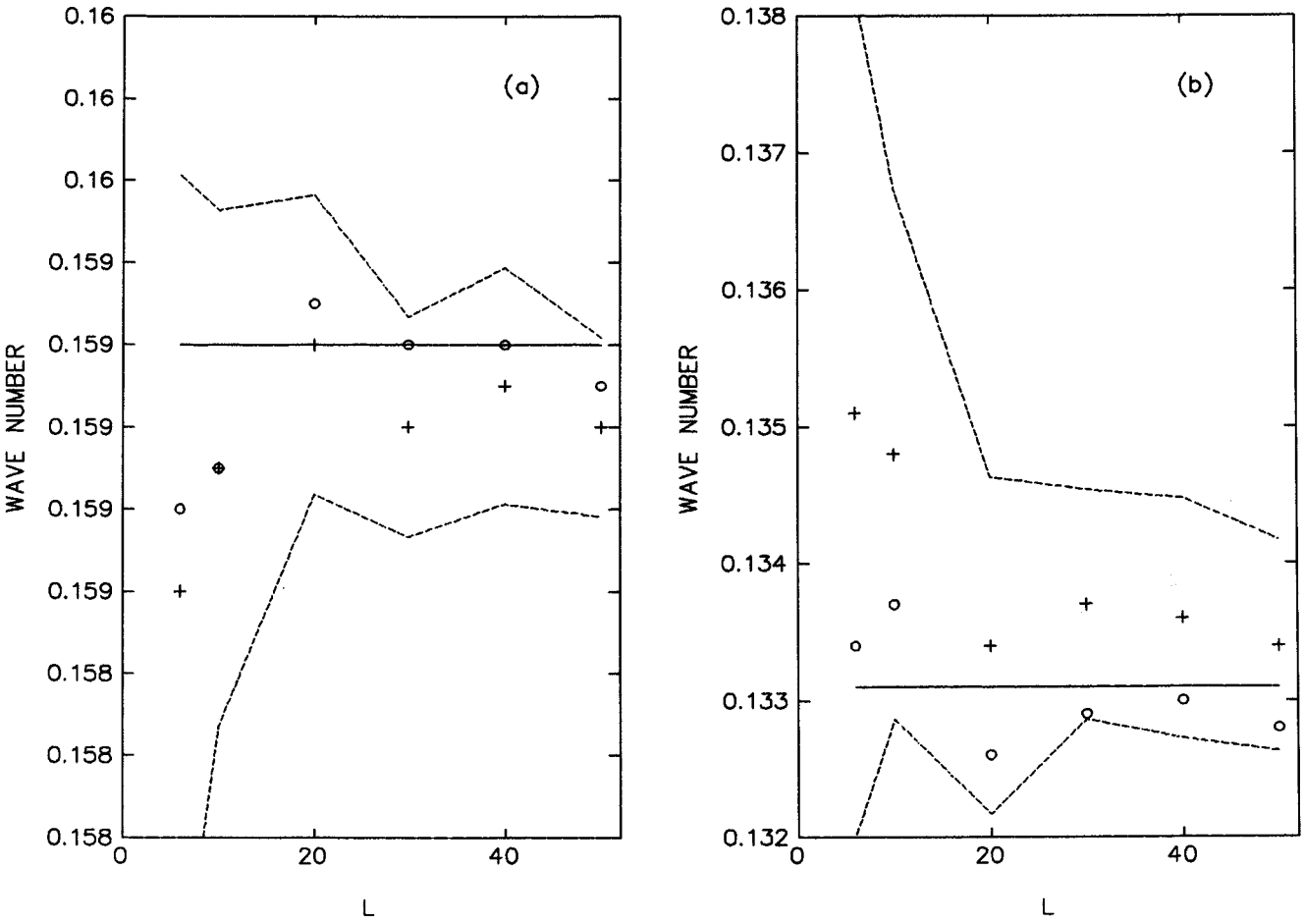


FIG. 4. Sampled means of the first (a) and the second (b) wave numbers as a function of L. Curve labels are as in Fig. 3.

In Fig. 3, the dependencies of the rms errors obtained for the first [Fig. 3(a), $k_1=0.1592$, $\tilde{g}_1=15.5$ dB] and the second [Fig. 3(b), $k_2=0.1331$, $\tilde{g}_2=5.5$ dB] wave numbers are plotted as functions of L. Analysis of these results shows that, beginning with the value $L\geqslant 10$ in both algorithms facilitates to achieve the limit wave-number estimation accuracy, i.e., CRB (the solid line in Fig. 3). When the number L of the observation vectors is sufficiently small, estimates k_m are no longer efficient, but, as earlier, the rms errors of both algorithms are approximately identical.

In Fig. 4, the sampled means of the first [Fig. 4(a)] and the second [Fig. 4(b)] wave numbers are plotted as functions of L. The horizontal lines relate to the true wave-number values. The estimates obviously remain unbiased in the broad range of values of L. Again, the WSF algorithm does not concede to the optimal algorithm.

Thus the computing experiment results show, that under restrictions on the accumulation time T_o, the WSF algorithm can compete with the optimal algorithm in the accuracy of the normal modes wave-number estimation.

VIII. SUMMARY AND CONCLUSIONS

This article is devoted to investigating of the structure and efficiency of normal mode wave-number estimation algorithms.

The synthesis and analysis of an optimal (by the criterion of the maximum likelihood function) algorithm for normal mode wave-number estimation are performed. The optimal algorithm has a high-resolution property, takes into account availability of correlated spatial noise, does not require prior modal decorrelation, is suitable for a horizontal array of any shape, and permits us to attain the potential wave-number estimation accuracy limit (Cramer–Rao bound).

As opposed to the MUSIC algorithm, the optimal algorithm eases estimation of the wave numbers of coherent (fully correlated) modes without their preliminary decorrelation. But if such decorrelation is performed, for example, by means of a moving array, then the optimal algorithm estimates the wave numbers with much higher accuracy (in terms of root-mean-square error) than the MUSIC algorithm.[9,10,19] Moreover, this difference in performance between the algorithms can be significant at small moving intervals when it is impossible to decorrelate the modes well enough.

Wave-number estimation, with the help of the optimal algorithm, requires a search for the global extremum of some goal function, with known computing difficulties and the absence of sufficiently well-developed computing algorithms. The algorithms based on the signal eigenvectors property of the spectral correlation matrix (SCM) measured on the array

elements are free from listed drawbacks. Synthesis of such algorithms has been accomplished here under the condition of arbitrary correlated spatial noise. As has been shown, among these algorithms there is one that permits us to receive accurate wave-number estimates such as those obtained by the optimal algorithm. This algorithm generalizes the known WSF algorithm in the case of spatially correlated noise.

As a result of the computing experiments, it has been established that the WSF algorithm can compete with the wave-number estimation accuracy with the optimal algorithm not only in the case of long time intervals used for the SCM estimation,[10,13,14] but also when this time interval is essentially limited.

ACKNOWLEDGMENT

The authors are grateful to the reviewers for the many useful comments made on earlier versions of this paper.

[1] W. Munk and C. Wunsh, ''Ocean acoustic tomography: A scheme for large scale monitoring,'' Deep-Sea Res. **26A**, 123–161 (1975).
[2] A. Tolstoy, O. Diachok, and L. N. Frazer, ''Acoustic tomography via matched field processing,'' J. Acoust. Soc. Am. **89**, 1119–2062 (1991).
[3] A. B. Baggeroer, W. A. Kuperman, and H. Schmidt, ''Matched field processing: Source localization in correlated noise an optimum parameter estimation problem,'' J. Acoust. Soc. Am. **83**, 571–587 (1988).
[4] E. C. Shang, Y. P. Wang, and Z. Y. Huang, ''Waveguide characterization and source localization in shallow water waveguides using the Prony method,'' J. Acoust. Soc. Am. **83**, 103–108 (1988).
[5] S. S. Reddi, ''Multiple source location—a digital approach,'' IEEE Trans. Aerosp. Electron. Syst. **AES-15**, 95–105 (1979).
[6] D. H. Johnson, ''The application of spectral estimation method to bearing estimation problems,'' Proc. IEEE **70**, 1018–1028 (1982).
[7] R. O. Schmidt, ''Multiple emitter location and signal parameter estimation,'' IEEE Trans. Antennas Propag. **AP-34**, 276–280 (1986).
[8] J. V. Candy and E. J. Sullivan, ''Model-based passive ranging,'' J. Acoust. Soc. Am. **85**, 2472–2480 (1989).
[9] P. Stoica and A. Nehorai, ''Performance study of conditional and unconditional direction-of-arrival estimation,'' IEEE Trans. Acoust. Speech Signal Process. **ASSP-38**, 1783–1795 (1990).
[10] B. E. Ottersten, ''Parametric subspace fitting methods for array signal processing,'' Ph.D. thesis, Stanford University, Stanford, CA (1989).
[11] H. M. Chouhan and G. V. Anand, ''Normal mode wave-number estimation using towed array,'' J. Acoust. Soc. Am. **93**, 1807–1814 (1993).
[12] J. A. Cadzow, ''Multiple source location—the signal subspace approach,'' IEEE Trans. Acoust. Speech Signal Process. **ASSP-38**, 1110–1125 (1990).
[13] M. Viberg and B. Ottersten, ''Sensor array processing based on subspace fitting,'' IEEE Trans. Signal Process. **39**, 1110–1121 (1991).
[14] M. Viberg, B. Ottersten, and T. Kailath, ''Detection and estimation in sensor arrays using weighted subspace fitting,'' IEEE. Trans. Signal Process. **SP-39**, 2436–2449 (1991).
[15] V. Kohler and G. C. Papanicolau, ''Wave propagation in randomly inhomogeneous ocean,'' in *Wave Propagation and Underwater Acoustics*, edited by J. B. Keller and J. S. Papadakis (Spring-Verlag, Berlin, 1977), pp. 126–179.
[16] S. Narasimhan and J. L. Krolik, ''Fundamental limits on acoustic source range estimation performance in uncertain ocean channels,'' J. Acoust. Soc. Am. **97**, 215–226 (1995).
[17] J. A. Cadzow, ''Direction-of-arrival estimation using signal subspace modeling,'' IEEE Trans. Aerosp. Electron. Syst. **AES-28**, 64–78 (1992).
[18] Amin G. Jaffer, ''Maximum Likelihood Direction Finding of Stochastic Sources: A separable solution,'' ICASSP **88**, 2893–2896 (1988).
[19] P. Stoica and A. Nehorai, ''MUSIC, Maximum Likelihood and Cramer-Rao Baund,'' IEEE Trans. Acoust. Speech Signal Process. **ASSP-37**, 720–741 (1989).
[20] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization* (Academic, London, 1981).
[21] I. Ziskind and M. Wax, ''Maximum likelihood localization of multiple sources by alternating projection,'' IEEE Trans. Acoust. Speech Signal Process. **ASSP-36**, 1553–1560 (1988).

Measurements and analysis of acoustic backscattering by elastic cubes and irregular polyhedra

Peter D. Thorne
Proudman Oceanographic Laboratory, Bidston Observatory Birkenhead, Merseyside L43 7RA, United Kingdom

Shaobo Sun
Department of Industrial Acoustics, Technical University of Denmark Building 425, DK-2800 Lyngby, Denmark

Jingdong Zhang
School of Physics, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom

Irina Bjorno
Department of Industrial Acoustics, Technical University of Denmark, Building 425, DK-2800 Lyngby, Denmark

Thierry Mazoyer
Metravid Research Development Service, 200 Chemin des Ormex, 69760 Limonest, France

(Received 22 August 1996; revised 30 May 1997; accepted 29 July 1997)

Underwater acoustic studies of backscattering by submerged targets have generally focused on bodies with spherical and cylindrical symmetry. However, there are interests in scattering by objects which may be characterized by more angular features, with surfaces that tend to be composed of facets and edges. To investigate the scattering properties of such bodies, the backscattering by a number of elastic cubes, and irregularly shaped polyhedra, have been studied. Data were collected by measuring the band limited impulse response of the scatterers, using a broadband transducer, which operated as a transceiver, both transmitting and receiving signals. To present the scattering measurements nondimensionally a form function definition has been employed to normalize the backscattered signals. The normalized frequency has been expressed as ka, where k is the acoustic wave number, and a is a characteristic dimension of the scatterer. The cube observations covered a broad ka range, $ka=3-34$, thereby encompassing the lower ka scattering region where the cube can be considered to be nominally rigid, through to higher ka values where it would be expected that the elastic properties of the cube become important. The measurements on irregular polyhedra are more limited in extent, however, the results are presented for comparison with the cube data. To compare the observations with predictions, computations of acoustic backscattering for both rigid and elastic cubes are presented. © *1997 Acoustical Society of America.* [S0001-4966(97)04811-X]

PACS numbers: 43.30.Gv, 43.30.Ma, 43.30.Ft

INTRODUCTION

Measurements and theoretical works on the interaction of underwater sound with elastic targets of spherical and cylindrical geometries have established a detailed understanding of the scattering properties of such bodies.[1–5] In particular the application of resonance scattering theory[6–10] has clarified the component structure of the scattered signal in terms of background and resonant components. While the scattering by canonically shaped targets is therefore well understood, the scattering by objects which lack spherical or cylindrical symmetry is less developed, and there are open questions regarding the scattering characteristics and equivalent form function for nonspherical targets. To investigate some of these questions, the present paper examines the backscattering by cubes, and irregularly shaped polyhedra.

Advancing our understanding of the scattering of underwater sound by objects which are not canonical, but are more angular in form, with features composed principally of edges and facets, has been stimulated in part because of its possible application to the interpretation of the interaction of sound with marine suspensions. In recent years acoustic backscatter systems[11–15] have been developed which can make detailed observations of near-bed suspended sediment processes. To extract suspended sediment parameters from the acoustic data entails an inversion procedure which is based on an assumed knowledge of the scattering characteristics of marine suspended sediments. To date the modeling of such particles has been based on spheres or fitting empirical algorithms to data sets.[16–18] Such pragmatic approaches have been very valuable, and they form the cornerstone of the acoustic inversion. However, as seen in Fig. 1, scanning electron micrographs show sediments composed of sand to contain a substantial proportion of particles with surfaces which have a tendency to be angular, and have a number of facets and edges. As a step toward identifying the scattering properties of such bodies, the backscattering characteristics of elastic cubes have been investigated in detail in the present work. Further, some limited measurements on irregular polyhedra are presented.

FIG. 1. Scanning electron micrography of quartz sand sediment particles.

To obtain the backscatter measurements a calibrated broadband transceiver was used, both transmitting the incidence signal toward the target, and receiving the backscattered signal from the target. The transducer transmitted a short wide-band pulse, and the band limited impulse response of the target was recorded. Data were collected on four different size cubes which covered a ka region of 3–34, and on three irregular polyhedra between $ka=7$ and 25.

Modeling the scattering characteristics of cubes and polyhedra is a relatively complex task because the wave equation is not separable for these geometries. Therefore analytical solutions are not available, and numerical methods are the only techniques presently able to predict the scattered field by such targets. In the present study theoretical approaches are employed which use the boundary element method (BEM) for rigid scatterers, and a combined BEM and finite element method (FEM) for elastic scatterers. Even with increasing access to powerful computing facilities, the computations are protracted, and the solutions tend to become more problematic at higher ka values. Therefore the calculations in the present work have been limited to upper values of $ka=20$ and $ka=16$, for the rigid and elastic cases, respectively.

I. NUMERICAL SCATTERING MODELS

In order to tackle the problem of scatterers in which the wave equation is not separable, numerical techniques remain the principal method for obtaining the scattered field. In the present study both the rigid and elastic response of a cube have been computed. To calculate the rigid response, a boundary element method has been applied to a rigid immovable cubic mesh in the frequency domain. One primary objective was to cover a broad ka range, and therefore particular care was taken to overcome the nonunicity of the standard BEM solution at certain characteristic frequencies. The solution employed was to use a Helmholtz gradient formulation (HGF), which consists of combining the Helmholtz integral equation with its normal derivative form.[19,20] Further, the utilization of variational formalism,[21] instead of the conventional collocation technique, leads to significant numerical simplification when computing the highly singular integrals involved in BEM. To compute the elastic response of a cube advantage was taken of a recently developed software package PAFEC.[22] PAFEC models the scattering problem using a combined FEM and BEM model. Finite elements are used to characterize the vibration of the elastic body, and boundary elements describe the radiation or scattering from the body. Even with the package, care was required in the formulation of the cube scattering problem, and an examination was needed to take advantage of available symmetries for particular cube scattering configurations, to optimize the intensive computations needed for the elastic case.

The cubes and polyhedra investigated were composed of iron. At the time of the experiments the material had been thought to be stainless steel, but later metallurgic analysis showed this not to be the case. This has introduced a degree of complication into the comparison between the observations and the computed elastic case because the elastic properties of iron tend to be less consistent than that of stainless steel. However, this said, useful comparisons have been made both with the experimental data and with the calculations for the rigid cube case. The parameters used for the elastic calculations were obtained from tabulated values for iron which gave compressional and shear wave velocities, respectively, of 5957 m s^{-1} and 3224 m s^{-1}, and a density of 7860 kg m^{-3}. The density and sound velocity of water used were respectively 1000 kg m^{-3} and 1460 m s^{-1}.

II. EXPERIMENTAL ARRANGEMENT

Measurements of the form function[23] were conducted in a small laboratory test tank with dimensions of $1.5\times0.5\times0.5$ m. The acoustic axis was directed along the center of the longest dimension of the test tank. A diagram of the instrumentation is shown in Fig. 2. The transducer used acted as a transceiver both transmitting and receiving signals. The transducer operated at a center frequency of 546 kHz and had an operating bandwidth between 300 and 800 kHz. Measurements were conducted by transmitting a single cycle tone burst at the center frequency, and the backscattered signal was gated, amplified, filtered, coherently averaged, digitized, and recorded. The voltage applied to the transducer was also digitized and recorded. The transmitting and receiving sensitivities of the transducer were measured using a reference hydrophone, and further self-reciprocity calibrations, and measurements[24] of the specular echo from a sphere confirmed the sensitivities. The transfer function of the electronics was also measured.

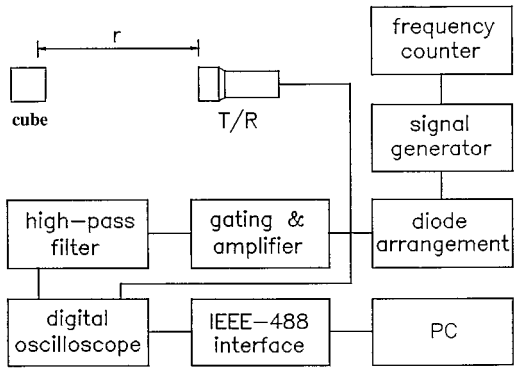


FIG. 2. Instrumentation used for the scattering measurements.

The experimental form function, f, for a three-dimensional scatterer can be defined as

$$f=\frac{2r}{a_{cf}}\frac{P_s(\nu)}{P_i(\nu)}, \tag{1}$$

where r is the distance between the center of the scatterer and the transducer, $P_s(\nu)$ is the spectrum of the scattered signal, $P_i(\nu)$ is the spectrum of the incident signal, ν is frequency, and a_{cf} is a characteristic dimension. In a recent study[25] it has been shown that the average projected area for any convex scatterer is a quarter of its surface area, and this leads to the correct value for a_{cf} of $a_{cf}=l(3/2\pi)^{1/2}$, where l is the side length of a cube. The appropriate value to normalize the frequency axis is less clear, however, comparisons of the form function for a rigid sphere, and a rigid cube[25] indicate that using the radius of a sphere having the same volume as the cube has advantages. This value is therefore given by $a_{ck}=l(3/4\pi)^{1/3}$, and the nondimensional frequency axis is given by ka_{ck}.

Re-expressing in terms of the measured signal f is given by

$$f=\frac{2r^2}{a_{cf}}\frac{V_s(\nu)}{V_i(\nu)G(\nu)S_{tr}(\nu)}, \tag{2}$$

where $V_s(\nu)$ is the spectrum of the backscattered voltage, $V_i(\nu)$ is the spectrum of the voltage applied to the transceiver, $G(\nu)$ is the gain of the electronics system, and $S_{tr}(\nu)$ is the transceiver sensitivity. To assess the validity of the form function measurements collected with the present system, observations of the scattering characteristic of a 12.5-mm-radius stainless steel sphere were recorded, and the form function evaluated using Eq. (2). Comparison with theoretical predictions are shown in Fig. 3. The similarity between the measured and predicted form function confirms the accuracy of the system calibration.

III. MEASUREMENTS AND ANALYSIS

Backscatter data were collected for varying angles of incidence and different cube configurations. Figure 4 shows the main configurations used, with the cube face-on, edge-on, and at an arbitrary configuration to the incoming signal. The cubes were rotated about the axis as shown and the variation in the backscattered form function with angle measured. The majority of data shown in the present work are associated with the face-on configuration [Fig. 4(a)], at varying angles of incidence, although to estimate the mean backscatter form function for a cube, measurements at different angles were also collected using the configurations shown in Fig. 4(b) and (c). The irregular shaped polyhedra data were collected by taking measurements at a number of angles of incidence (simply referenced to a starting position and rotated), and configurations.

Measurements of the backscattered signal from a cube in the time domain are shown in Fig. 5. To illustrate the form and duration of the signal employed Fig. 5(a) shows the specular component of the backscattered signal from a 12.5-mm-radius sphere. Figure 5(b)–(d) shows the backscattered signal for the face-on configuration at $\theta=0°$ (normal incidence), 10°, and 45°, for a cube with 8-mm-length sides. The 0° shows a time series echo that is comparable to the specular echo from the sphere, with the signal being primarily composed of a simple single pulse. By 10° the amplitude of the backscattered signal has reduced by about a factor of 5,

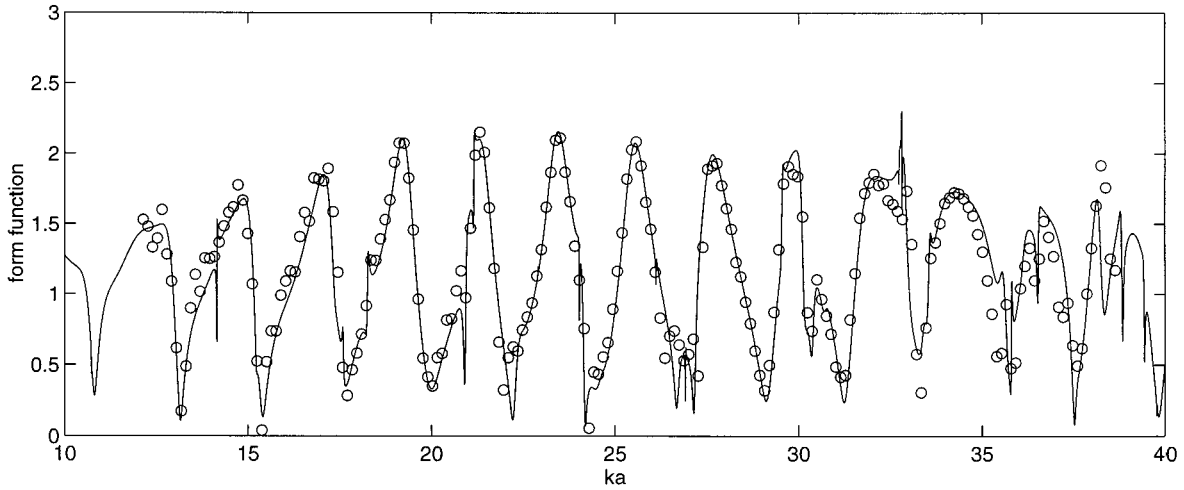


FIG. 3. Comparison of the measured (○) and the predicted (———) form function for a 12.5-mm-radius stainless steel sphere.

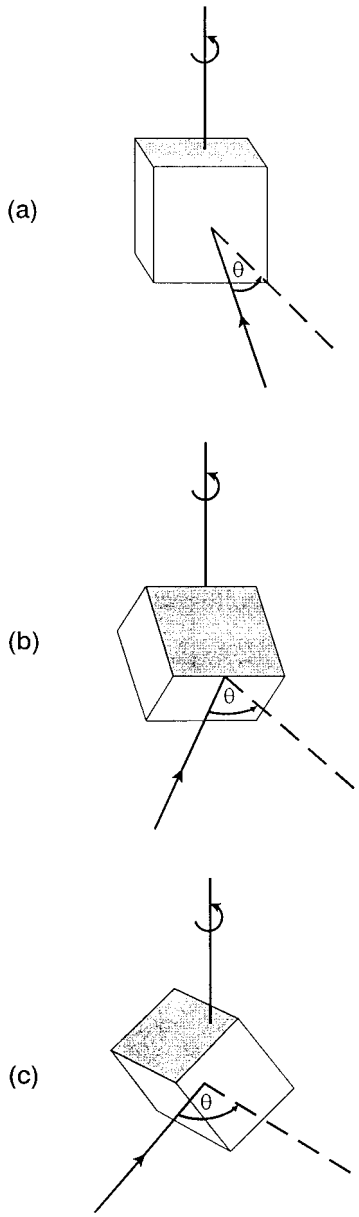


FIG. 4. Various configurations of the cube mounting. (a) Face-on; (b) edge-on; and (c) arbitrary configuration.

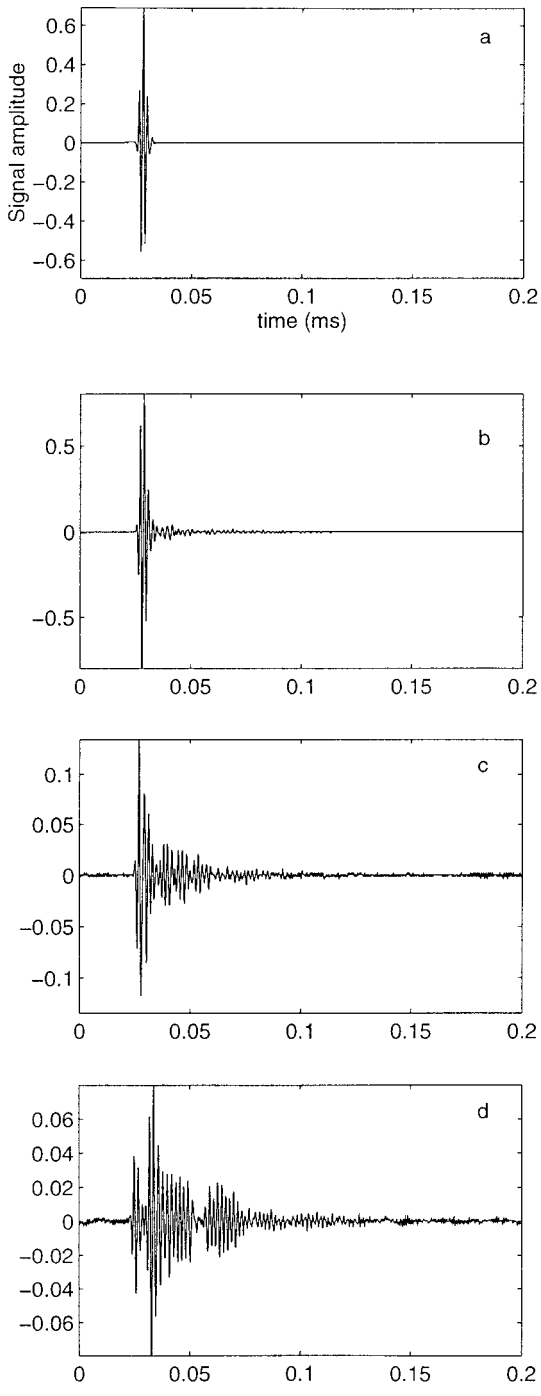


FIG. 5. Backscattered time series measurements. (a) Specular echo from a 12.5-mm-radius sphere. Backscattered echo for the face-on configuration for a cube at: (b) $\theta=0°$; (c) $\theta=10°$; and (d) $\theta=45°$.

and there is a structured echo with a duration which is significantly greater than that for 0°. By 45° (which is the same as the normal incidence edge-on case) the signal amplitude is now around one-tenth of the $\theta=0°$ case. The signal has an initial arrival which is considered to be due to the leading edge, and this is followed by the bulk of the signal. The latter is relatively complex in structure and has a duration which is somewhat longer than the 10° result.

From the time domain waveforms the backscattered form functions have been computed using Eq. (2). The normal incidence results, $\theta=0$, for the face-on configuration are shown in Fig. 6. The data were collected on cubes with side lengths of 4 (+), 8 (○), 12 (*), and 16 (×) mm. Unlike sphere scattering, which results in a form function which oscillates about unity with increasing ka, owing to its flat surface, the form function for each cube is observed to steadily increase with ka. Superimposed upon this general trend are a series of dips, and the ka location of these dips are consistent for the different cubes. Two theoretical form functions are compared with the data in Fig. 6(a), the rigid immobile cube (-.-), and the elastic cube (—). It can be seen that the rigid case compares well with the data in the low ka_{ck} region. This is as might be expected for an iron cube, since a rigid model should be a reasonable approximation at low ka_{ck} values. Above a $ka_{ck}\approx 6$ the rigid predictions are in agreement with the general trend of the data although the dips in the form function are not replicated. The elastic re-

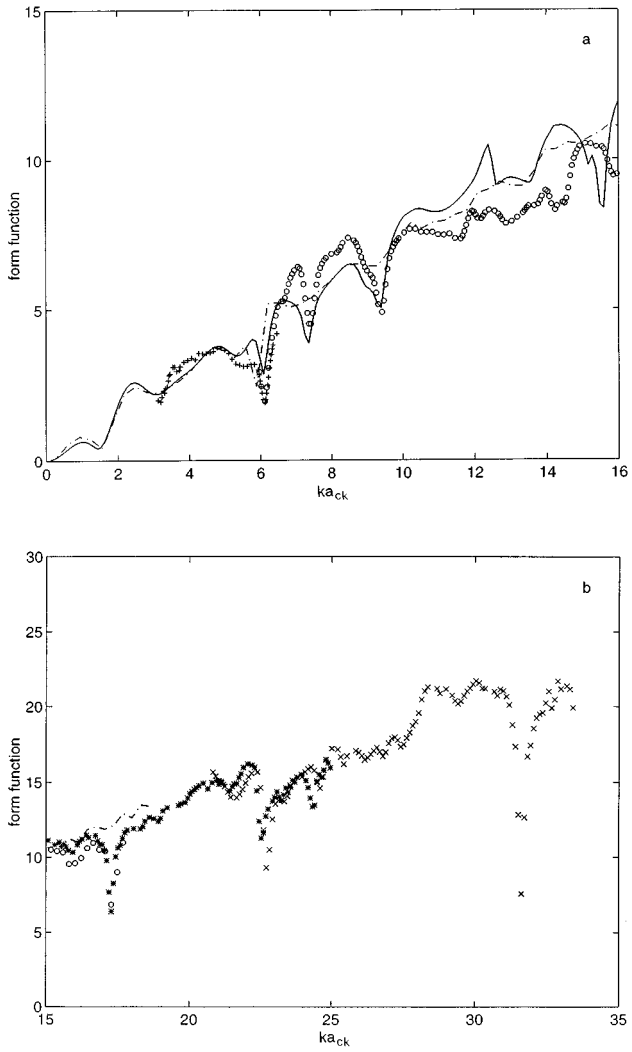


FIG. 6. Comparison of the measured and the predicted form function at $\theta = 0°$ for the face-on configuration. Measurements were taken using 4 (+), 8 (○), 12 (*), and 16 (×) mm sized cubes. Predictions were computed using rigid (-·-) and elastic (———) models for a cube.

sults, however, show both general agreement with the data, and the dips in the form function at $ka_{ck}=6$, 7.4, and 9.4 are replicated. There does appear to be some discrepancy between predictions and observations just below $ka_{ck}=6$, and for $ka_{ck}>10$. This could be associated with the uncertainty in the estimated values for the compressional and shear wave velocity of the iron used to form the cubes. However, because of the extensive computations required to obtain the elastic results, up to periods of several days, it is presently impractical to run a series of scenarios with varying sound speeds to investigate the details of the difference. Also the precision of the alignment, considered to be of the order of 1° in both the rotational and vertical direction, may be influencing the level of the higher ka values. Notwithstanding these limited differences between theory and experiment, the main features of the normal incidence data are well represented using the elastic formulation for cube scattering. In Fig. 6(b) results are presented up to $k_{ck}=33$. These continue to show an increase in form function with k_{ck} with a number of dips. There are no elastic calculations for this regime,

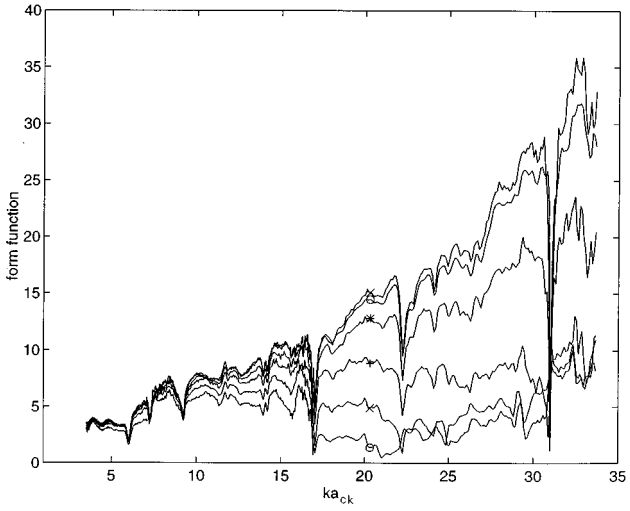


FIG. 7. Variation in the form function at near-normal incidence angles for the face-on configuration. x (upper curve)—$\theta=0°$, ○—$\theta=1°$, *—$\theta=2°$, +—$\theta=3°$, ×—$\theta=4°$, and ○ (lower curve)—$\theta=5°$.

although values for the rigid case are presented up to $k_{ck}=20$. Again the rigid case provides the general form but not the notable dips observed in the data. Since from Fig. 6(a) these dips are replicated in the elastic case, it would appear that dips are associated with modes of vibration of the cube which interfere destructively with the sound reflected from the surface of the cube, the latter being represented by the rigid calculation. To investigate the angular stability of the form function dips observed in the theory and the experiment at $\theta=0$, measurements were obtained off normal incidence. Figure 7 shows the outcome from some of these observations for $\theta=0°-5°$. It can be clearly seen that although the amplitude of the form function reduces rapidly off normal incidence, there is coherence in the location of the dips out to about 5°.

Figures 8–11, respectively, show measurements of the backscattered form function at 5°, 10°, 30°, and 45° for the

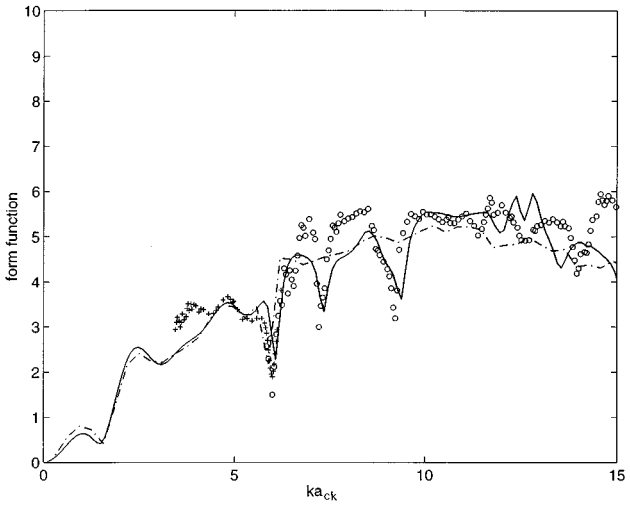


FIG. 8. Comparison for the face-on configuration at $\theta=5°$ of the form function computed for rigid (-·-) and elastic (———) cubes, and measured using 4 (+) and 8 mm (○) sided cubes.

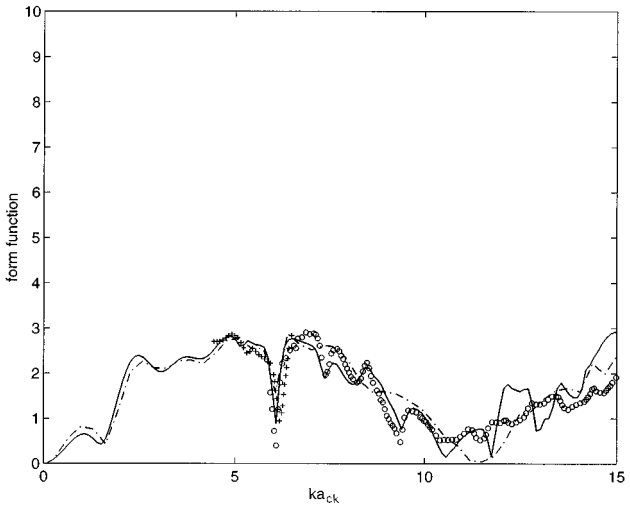


FIG. 9. Comparison for the face-on configuration at $\theta=10°$ of the form function computed for rigid (-·-) and elastic (———) cubes, and measured values using 4 (+) and 8 mm (○) sided cubes.

face-on configuration. The experimental data shown in Fig. 8 have form function values nominally between four and seven, with a series of identifiable dips, which as noted above, are present for all near-normal angles of incidence. For the 5° case the rigid computations provide a reasonable estimate for the general shape of the form function, however, the detailed variability associated with the cube being elastic is not reproduced. The elastic case provides results which are comparable with the observation, and with the dips in the form function being present. The results shown in Fig. 9 for a 10° angle of incidence show a different structure to those at 5°, with a reduction in the general level of the form function near $ka=10-12$. Predictions based on rigid and elastic cubes are compared with the data. The rigid results again provide the general form of the data, while it requires the elastic solution to replicate the detailed variability in the observations, although beyond $ka_{ck}=10$ there are differences between prediction and observation, these are probably due

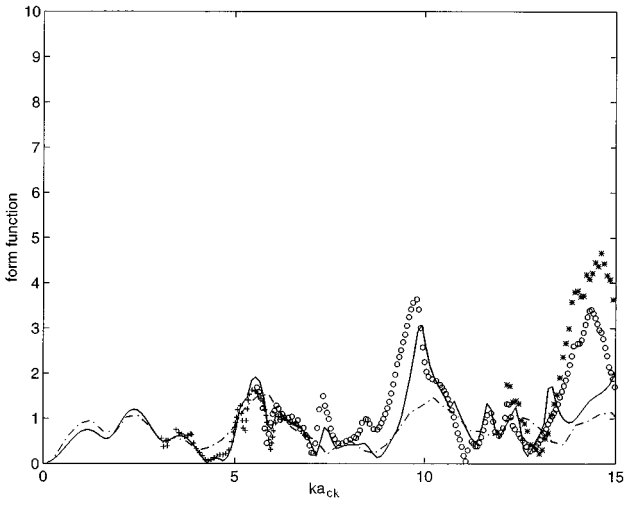


FIG. 10. Comparison for the face-on configuration at $\theta=30°$ of the form function computed for rigid (-·-) and elastic (———) cubes, and measured using 4 (+), 8 (○), and 12 mm (*) sided cubes.

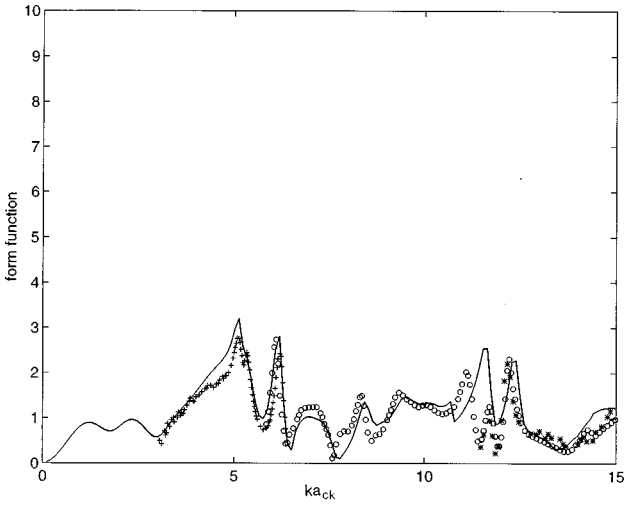


FIG. 11. Comparison for the face-on configuration at $\theta=45°$ (equivalent to the edge-on configuration at $\theta=0°$) of the form function computed for an elastic cube (———), and measured values using 4 mm (+), 8 mm (○), and 12 (*) mm sided cubes.

to uncertainties in the compressional and shear wave velocities for iron. The 30° results are shown in Fig. 10. The structure of the form function is quite variable with a series of peaks and troughs in the data. The general background trend with a number of dips superimposed upon it observed at the near-normal angles is not seen, and the form function is generally more complex. To compare the observations with the predictions, computation for a rigid and an elastic cube were carried out. There is seen to be reasonable agreement between the rigid calculations and measurements for $ka\leqslant 7$, however, at higher ka values the comparison between the rigid computations and the data is poor due to the onset of the elastic characteristics of the cube. The elastic computations show significantly improved agreement with the data, with, for example, the peak in the form function near $ka_{ck}=10$ being replicated, albeit at a slightly higher ka_{ck} value. Figure 11 shows measurements and computed predictions for the elastic case at $\theta=45°$; this configuration is equivalent to the configuration in Fig. 4(b) at $\theta=0°$. The data show notable peaks in the form function around $ka_{ck}\approx 5$ and $ka_{ck}\approx 6$, followed by a form function structure which is variable in form. The elastic calculations are seen to be in reasonable agreement with the data for $ka_{ck}=0-15$ for this angle. The fact that for some angles the comparison between the elastic and measured data are better than at other angles possibly indicates a degree of anisotropy in the material, this would not necessarily be unexpected for iron. At present for the elastic case computational periods are too extensive to iterate the compressional and shear wave velocities, and optimize the comparison at each angle.

The form function comparisons at particular angles allow the theory and experimental results to be examined in detail, however, they do not provide a global view of the scattering characteristics for a cube. To present a more general impression of the scattering properties, 3-D plots of the form function for 4-, 8-, and 16-mm-sided cubes are shown in Fig. 12(a)–(c). These were obtained using the face-on configuration, and the cube rotated from $\theta=0°$ to 45° in 1°

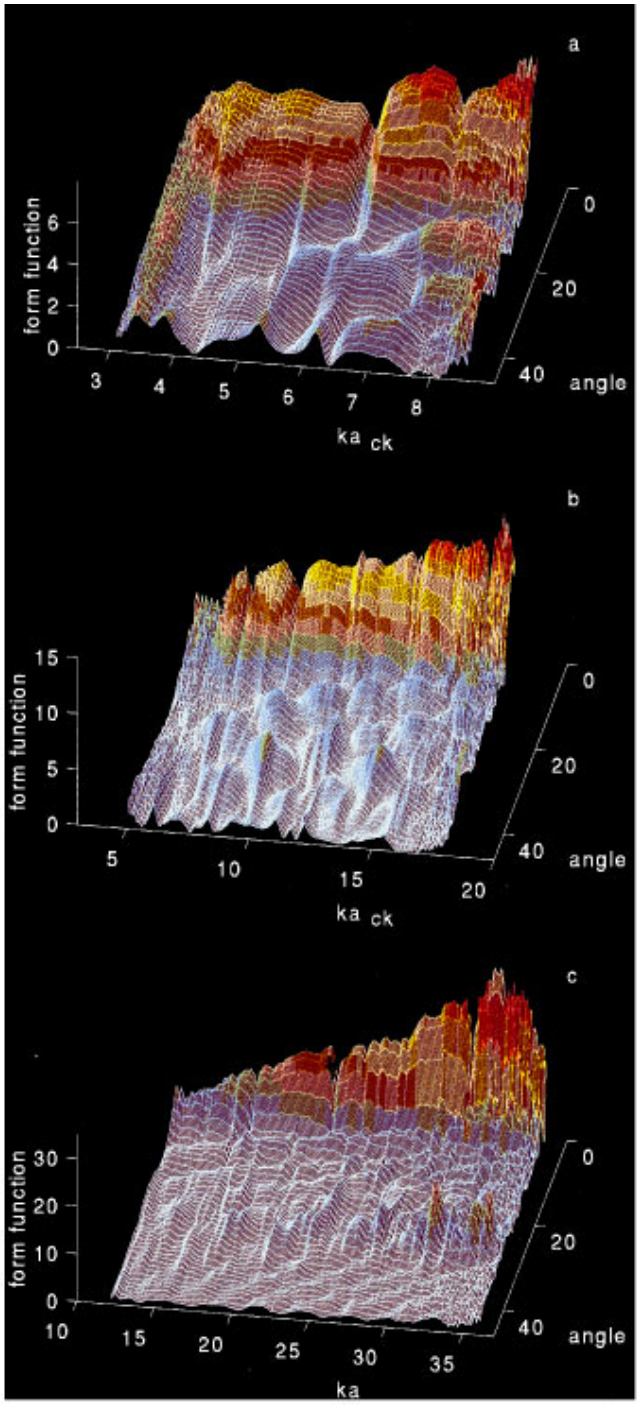


FIG. 12. Measurements of the variation of the form function for the face-on configuration with ka_{ck} and θ for (a) 4, (b) 8, and (c) 16 mm sided cubes.

steps. Near normal incidence each plot clearly shows high form function values with readily identifiable dips. These high values rapidly reduce away from normal incidence, and beyond about 10° the form function has an undulating structure both in θ and ka_{ck}, with values which typically oscillate between approximately 0 and 2. The physical interpretation of such figures is difficult at present because of the lack of an analytical representation of the cube scattering, and the long periods required to compute the results. However, the plots do provide a broad picture of cube scattering for the face-on configuration.

In part, the stimulation for developing a description of the scattering properties of a cube was an interest in the scattering properties of suspensions of marine sediments. It was therefore considered useful to examine the mean scattering properties of a cube. This was carried by taking a number of measurements using the three configurations shown in Fig. 4. A total of 107 form function measurements were taken on an 8-mm-side cube, and the average result from this data set is shown in Fig. 13. This result shows an average form function for the data set which is nominally centered about unity, with a number of notable peaks between $ka_{ck}=6$ and 10, and a steady oscillation in the form function for $ka_{ck}>10$. The peaks in the $ka_{ck}=6-10$ region do appear to be genuine features; they were noticeable in the individual form functions for some of the arbitrary configurations. The regular oscillations above $ka_{ck}=10$ are close in amplitude to the normalized error in the mean, which for 107 measurements was 9.7%; however, the regularity does suggest it may well be an actual feature. Two curves are compared with the data, i.e., the form function for a rigid sphere (⋯), and the average form function for a rigid cube (-.-). It is interesting to note that the theoretical form functions are comparable, and in particular the feature of a diffraction type oscillation for the average form function for a cube below $ka_{ck}=5$ is readily noticeable. Recent measurements[25] on a cube in this low ka_{ck} regime have clearly identified this structure. The rigid cube calculations, which were computed for 36 configurations, give approximately the same level as the observations, although the peaks are not observed and the higher ka_{ck} oscillations are not readily decernable. The former is ascribed to the model being rigid, while the latter may be due to the relatively coarse value for the ka_{ck} interval which was 0.31.

The irregular polyhedra were formed by taking iron cubes, and filing the edges until the shape was comparable to the sand particles shown in Fig. 1. Both the surface area and volume of the irregular polyhedra were measured to appropriately normalize the data. Measurements were taken on three irregular polyhedra having a_{ck} values of 4, 6.3, and 8.1 mm. Data were obtained for three configurations, and eight different angles of incidence for each polyhedron, and the mean form function for each polyhedron calculated. The average form functions was obtained by taking the average of the three mean form functions. The final result is presented in Fig. 14 and this shows the irregular polyhedra to have an average form function which is nominally uniform, and with a magnitude close to unity. Two curves are compared with the data, those of a rigid sphere (⋯) and a rigid icosahedron (-.-). The icosahedron was chosen because the irregular polyhedra had of the order of 20 sides, and its symmetry was amenable to computation. The computational results for the icosahedron are seen to be very similar to those of a rigid sphere with the diffraction oscillation clearly delineated. The rigid sphere could be thought of as a rigid polyhedron with an infinite number of facets; therefore as the number of facets on a polyhedron increase it might be expected that the solution would tend toward that of a sphere. Comparison of the computations with the observations show good agree-

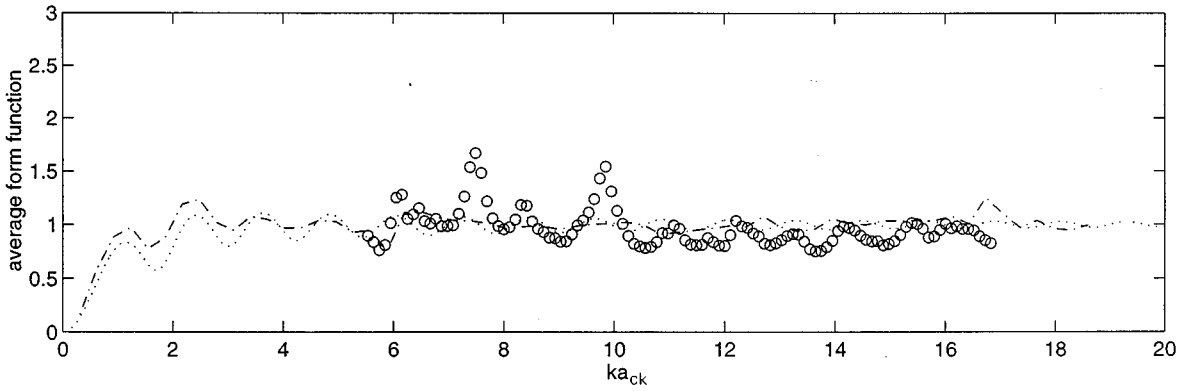


FIG. 13. Comparison of the average form function computed for a rigid cube (-·-) and a rigid sphere (···), with measured values for a 8 mm (○) sided cube.

ment between the predicted and the measured average form functions.

IV. CONCLUSIONS

A series of measurements on the backscattering by submerged elastic cubes have been conducted, and the results formulated nondimensionally using the form function definition, with a characteristic dimension defined by the projected surface area. The observations show a steadily increasing form function with ka_{ck} for normal incidence with the face-on configuration. Rotation off normal incidence results in a rapid reduction in the amplitude of the form function, although there is a degree of coherence in the structure of the form function up to about 5°. Beyond about 5° the composition of the form function becomes more complex, especially at the higher ka_{ck} values when the elastic nature of the cube becomes important. To provide a global sense of these relatively complex scattering characteristics, 3-D plots have been generated, and these provide some impression of the overall $ka_{ck}-\theta$ structure.

To theoretically describe the scattering properties of a cube, numerical computations using both rigid and elastic models have been evaluated. The rigid predictions provide solutions which are valid at low ka_{ck}, and give a first approximation to the general structure of the form function over a broad ka_{ck} range at near-normal incidence for the face-on configuration. To describe the detailed variability in the form function, and off normal incidence angles requires the elastic model to be run, and this does indeed provide significantly improved agreement between the predictions and the observations. Even with the elastic model there are discrepancies between computed values and measured data, and these are considered to be associated with uncertainties in the appropriate values for the compressional and shear wave velocities for iron.

Finally configuration and angular averaged form functions have been presented for a cube and irregular polyhedra. These show form functions which nominally have amplitudes of the order of unity, although in the cube case a persistent regular structure is observed. A number of predictions have been made for the average form function response, with rigid sphere, cube, and icosahedron shapes being computed. An interesting feature of the computations is the diffraction type oscillations observed in the cube and icosahedron form functions in the low ka_{ck} region which appear comparable in periodicity to those on a sphere generated by creeping waves. Such oscillations have been recently observed experimentally[25] for a cube in the $ka_{ck}=0-5$ region, and thereby validate the predictions. It therefore appears that there could be some kind of specular, and creeping wave interaction which is emerging through the mean form function that is comparable with that of a sphere. Comparison of

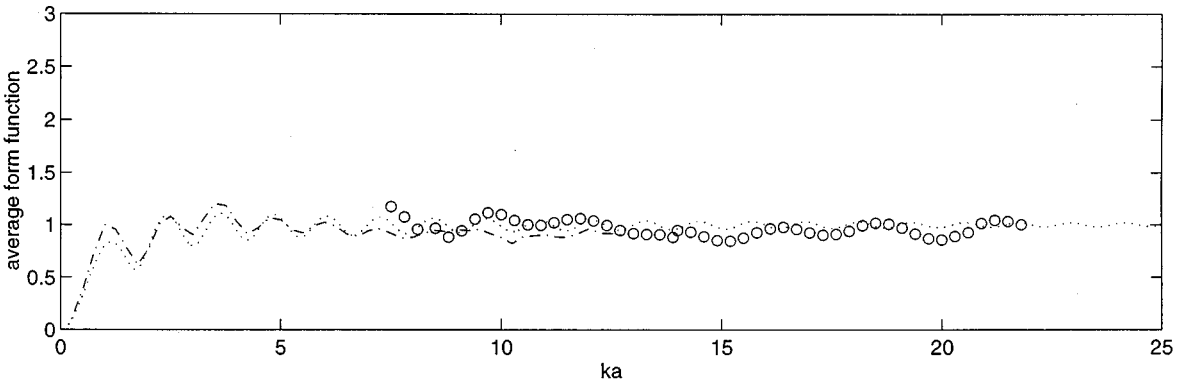


FIG. 14. Comparison of the average form function computed for a rigid icosahedron (-·-) and a rigid sphere (···), with measured values (○) obtained using irregular polyhedra with a_{ck} values of 4, 6.3, and 8.1 mm.

the computed form functions with the data show reasonably consistent results, and the results indicate why the sphere based scattering models for sediment scattering yield reasonable results. To first order the mean form function for rigid polyhedra, (which from Fig. 1 would appear to be a more appropriate description for sandy sediments) are similar to that of a rigid sphere.

ACKNOWLEDGMENTS

This work was supported by the European community under its MAST programs, and by the coastal protection division of MAFF, UK. P.D.T. became involved in this work during a visit to the Department of Industrial Acoustics at the Technical University of Denmark, and he would like to thank Professor Leif Bjorno for his invitation and support. Thanks also go to Dr. Victor Humphrey and Dr. Paul Chinnery, of Bath University, U.K., for a number of interesting discussions on scattering problems.

[1] J. J. Faran, ''Sound scattering by solid cylinders and spheres,'' J. Acoust. Soc. Am. **23**, 405–418 (1951).
[2] W. G. Neubauer, R. H. Vogt, and L. R. Dragonette, ''Acoustic reflections from elastic spheres. I. Steady-state signals,'' J. Acoust. Soc. Am. **55**, 1123–1129 (1974).
[3] P. L. Marston, ''GTD for backscattering from elastic spheres and cylinders in water and the coupling of surface elastic waves with the acoustic field,'' J. Acoust. Soc. Am. **83**, 25–37 (1988).
[4] R. Hickling, R. K. Burrows, and J. F. Ball, ''Rotational waves in the elastic response of spherical and cylindrical acoustic targets in water,'' J. Acoust. Soc. Am. **89**, 971–979 (1991).
[5] R. H. Hackman, G. S. Sammelmann, K. L. Williams, and D. H. Tivett, ''A reanalysis of the acoustic scattering from elastic spheroids,'' J. Acoust. Soc. Am. **83**, 1255–1266 (1988).
[6] L. Flax, L. R. Dragonette, and H. Uberall, ''Theory of elastic resonance excitation by sound scattering,'' J. Acoust. Soc. Am. **63**, 723–731 (1978).
[7] G. C. Gaunaurd and H. Uberall, ''RST analysis of monostatic and bistatic acoustic echoes from an elastic sphere,'' J. Acoust. Soc. Am. **73**, 1–12 (1983).
[8] W. G. Neubauer, ''Acoustic reflection from surfaces and shapes,'' published by Naval Research Laboratory, Washington, DC (1986).
[9] M. F. Werby and G. C. Gaunard, ''Resonance scattering from submerged elastic spheroids of high aspect ratios and its three-dimensional interpretation,'' J. Acoust. Soc. Am. **88**, 951–960 (1990).
[10] D. Decultot, F. Lecroq, G. Maze, and J. Ripoche, ''Acoustic scattering from a cylindrical shell bounded by hemispherical end caps. Resonance interpretation with surface waves propagating in cylindrical and spherical shells,'' J. Acoust. Soc. Am. **94**, 2916–2923 (1993).
[11] A. E. Hay and J. Sheng, ''Vertical profiles of suspended sand concentration and size from multifrequency acoustic backscatter,'' J. Geophys. Res. **97**, 15 661–15 677 (1992).
[12] P. D. Thorne, P. J. Hardcastle, and R. L. Soulsby, ''Analysis of acoustic measurements of suspended sediments,'' J. Geophys. Res. **98**, 899–910 (1993).
[13] P. D. Thorne, P. J. Hardcastle, and A. Hogg, ''Observations of near-bed suspended sediment turbulence structures using multifrequency acoustic backscattering,'' in *Coherent Flow Structures on Open Channels*, edited by P. J. Ashworth, S. J. Bennett, J. L. Best, and S. J. McLelland (Wiley, London, 1996), pp. 343–358.
[14] J. F. Lynch, J. D. Irish, C. R. Sherwood, and Y. C. Agrawal, ''Determining suspended sediment particle size information from acoustical and optical backscatter measurements,'' Continental Shelf Res. **14**, 1139–1165 (1994).
[15] J. Sheng and A. E. Hay, ''Sediment eddy diffusivities in the nearshore zone, from multifrequency acoustic backscatter,'' Continental Shelf Res. **15**, 129–147 (1995).
[16] J. Sheng and A. E. Hay, ''An examination of the spherical scatterer approximation in aqueous suspensions of sand,'' J. Acoust. Soc. Am. **83**, 598–610 (1988).
[17] A. M. Crawford and A. E. Hay, ''Determining suspended sand size and concentration from multifrequency acoustic backscatter,'' J. Acoust. Soc. Am. **94**, 3312–3324 (1993).
[18] P. D. Thorne, K. R. Waters, and T. J. Brudner, ''Acoustic measurements of scattering by objects of irregular shape,'' J. Acoust. Soc. Am. **97**, 242–251 (1995).
[19] C. M. Piaszczyk and J. M. Klosner, ''Acoustic radiation from vibrating surfaces at characteristic frequencies,'' J. Acoust. Soc. Am. **75**, 363–375 (1984).
[20] Z. Ruet, ''On the boundary integral methods for the exterior acoustic problem,'' J. Sound Vib. **103**, 297–298 (1985).
[21] M. Mamdi, Ph.D. thesis, University de technology de Compiegne, France, 1982.
[22] PAFEC VibroAcoustics, from Pafec Ltd, Strelley Hall, Nottingham N68 6PE, U.K.
[23] S. Sun, P. D. Thorne, I. K. Bjørnø, and T. Mazoyer, ''Observations of acoustic backscattering by elastic cubes,'' in The Proceedings of the 'Third European Conference on Underwater Acoustics' edited by J. S. Papadakis, and held at FORTH/IAMC on Crete, 24–28 June 1996, pp. 51–56.
[24] S. Sun, P. D. Thorne, and I. K. Bjørnø, ''Using the specular component of sphere backscattering for broadband transceiver calibration,'' Acust. Acta. Acust. **82**(4), 672–674 (1996).
[25] P. A. Chinnery, V. F. Humphrey, and J. Zhang, ''Low frequency acoustic backscattering by a cube: Experimental measurements and theoretical prediction,'' J. Acoust. Soc. Am. **101**, 2571–2582 (1997).

Thermoacoustic relaxation in a pin-array stack

M. E. Hayden and G. W. Swift
Condensed Matter and Thermal Physics Group, Los Alamos National Laboratory, MS K764, Los Alamos, New Mexico 87545

(Received 2 April 1997; accepted for publication 9 July 1997)

The pressure response p_1 of a fixed quantity of ^{4}He gas subjected to sinusoidal volume perturbations is studied experimentally. Thermal diffusion near solid boundaries causes the gas response to be neither adiabatic nor isothermal. The data are interpreted in terms of the thermal relaxation function f_κ which relates p_1 to the spatial average of the acoustic temperature profile. Comparison of a number of geometries allows us to accurately determine f_κ for a thermoacoustic pin-array stack. [S0001-4966(97)00611-5]

PACS numbers: 43.35.Ud, 43.58.Bh [HEB]

INTRODUCTION

Thermoacoustic processes comprise the interactions between the displacement, pressure, density, and temperature oscillations of an acoustic wave in a compressible gas or other fluid, which are manifest near a solid, immobile surface acting as a thermal reservoir. These interactions occur within a boundary layer whose thickness is of order the characteristic length scale for thermal diffusion during one period of an acoustic cycle. The fact that the temperature of the fluid is coupled to that of the solid at the acoustic frequency, yet only moderately so, results in oscillations which are neither adiabatic nor isothermal. Phase shifts introduced by this coupling can lead to a variety of heat pumping effects.[1] Over the course of the past several decades, these phenomena have been harnessed to create a new class of powerful heat engines and refrigerators.[2]

It is convenient to use the thermal penetration depth

$$\delta_\kappa = \sqrt{\frac{2\kappa}{\omega}} \tag{1}$$

as a definite measure of the length scale for oscillatory thermal diffusion in a gas with thermal diffusivity κ. The thermal penetration depth is approximately the distance that heat will diffuse in a time $2/\omega$, where ω is the angular frequency of the acoustic oscillations under consideration. Gas situated much farther than δ_κ from a solid body cannot communicate thermally with that body and thus responds adiabatically to an acoustic wave. Very close to the gas–solid interface, however, thermal diffusion occurs rapidly and the thermal inertia of the solid forces the gas to respond isothermally. More complex thermoacoustic processes occur at intermediate distances, of order δ_κ away from the solid, where there is significant thermal coupling between the two media yet the time required for the transfer of heat is of order half the acoustic period.

The crucial element of many thermoacoustic devices is a passive structure generally referred to as a ''stack.'' This can be any porous medium with substantial surface area such that most of the entrained gas lies within the boundary layer, and yet which does not significantly impede the acoustic wave.[3] In effect, the stack is no more than a means for increasing the total volume of gas in which heat pumping effects can occur. Acoustic power densities greater than 10^5 W/m^3 are often generated or dissipated in the stacks of modern thermoacoustic heat engines and refrigerators.

Practical thermoacoustic devices employ a variety of stack geometries ranging from arrays of parallel-plate channels to hexagonal pores.[4,5] Theoretical arguments[6] indicate that a pin-array stack[7] (a triangular array of rods or ''pins'' aligned with the acoustic wave vector) should be more efficient than other geometries. A simple explanation of why this should be so is related to the fact that thermoacoustic heat pumping effects tend to occur further from the solid than the region of maximum viscous shear (or power dissipation).[8] Highly curved surfaces can modify the relative importance of these two effects with respect to a planar geometry. A convex surface (e.g., that of a pin) accentuates useful thermoacoustic processes relative to the viscous dissipation of power. A concave surface (e.g., that of a pore), on the other hand, diminishes thermoacoustic efficiency relative to viscous processes. Experimental results[9] indeed appear to confirm the prediction that the pin-array geometry is superior to other stack geometries.

Theoretical analyses of this type of problem date back to Kirchhoff's investigation of sound propagation in channels with circular cross section.[10] More recently, and primarily in the context of thermoacoustic devices, a number of authors have examined a range of geometries[11,12] including pin arrays.[6,13,14] One finds that the transverse spatial average of the complex amplitude of the oscillating component of the temperature T_1 is related to the complex acoustic pressure amplitude p_1 through a complex function f_κ of the thermal penetration depth and the geometry of the stack such that

$$\frac{\langle T_1 \rangle}{T_m} = \left(\frac{\gamma - 1}{\gamma} \right) \frac{p_1}{p_m} (1 - f_\kappa) \tag{2}$$

for an ideal gas. Here T_m and p_m are the mean temperature and pressure of the gas, γ is the ratio of the isobaric to isochoric specific heats, and the $\langle \cdots \rangle$ denote the transverse spatial average. We note that the complex function f_κ takes on the values 0 and 1 in the adiabatic and isothermal limits, respectively. The significance of this function is that, in effect, it characterizes the thermoacoustic performance of a stack or other acoustic element.[1,12]

In this paper, we present a direct measurement of f_κ for a thermoacoustic pin-array stack. The motivation for this work is twofold. First, a practical realization of the pin-array stack geometry is a significant engineering challenge[6,9] because the length scale δ_κ is typically only a small fraction of a millimeter in most thermoacoustic devices.[2] Here, we take advantage of the principles of similitude[15] to construct a simple ''large-scale'' model in which all of the relevant dimensions are much larger. In this manner, a well defined and characterized geometry is available for study. The results we obtain are directly applicable to practical pin-array stack designs. Second, existing derivations of f_κ for a pin-array stack[6,13,14] make a simplifying assumption regarding the symmetry boundary condition between adjacent hexagonal unit cells of the array. In particular, each hexagonal unit cell around a pin is assumed to be equivalent to a circle of equal area on whose perimeter radial gradients of T_1 are zero. It is prudent to evaluate this approximation before proceeding with future practical designs.

The work we describe is similar to previous studies of sound propagation in single pores[17] and pore arrays.[18–20] The principle innovation we introduce is a method for treating the influence of finite-size effects. This allows one to examine complex structures such as the pin-array stack for which theoretical calculations are difficult. A significant difference from earlier studies, related to our decision to use a similitude-correct large-scale model of a stack, is that our experimental chamber is small enough compared to an acoustic wavelength that it can be considered spatially isobaric.

I. THEORY OF THE MEASUREMENT

In the present work we employ a difference technique to determine f_κ for a thermoacoustic pin-array stack. Generalizations of the idea to other geometries should be straightforward. Consider, as an example, the two enclosed volumes shown in Fig. 1. We assume that they have been filled to the same mean pressure with an ideal gas and that they are identical in all respects except that one houses the pin-array stack under investigation. For convenience, we refer to these two cases as the ''empty'' and ''full'' configurations.

As shown in the figure, the total volume V_{tot} of the full container has been partitioned such that the internal volume V_i is that volume which includes the pin-array stack. The remaining volume V_e is exterior to the stack. The exact location of the boundary between V_i and V_e is somewhat vague in this example as it obviously depends upon δ_κ; this issue will be clarified in Sec. III. Regardless, V_i can be further divided into the volume V_{ip} occupied by the pins (or other solid material) and that occupied by the gas V_{ig}. Finally, the volume V_{tot} of the empty container has been partitioned into two volumes V_i and V_e such that they are identical to those of the full configuration.

If the volume of the containers is perturbed sinusoidally at angular frequency ω, pressure oscillations will be induced in the gas. We assume that the volume changes with time according to

$$V(t) = V_{\text{tot}} + \text{Re}[V_1 \exp(i\omega t)] \tag{3}$$

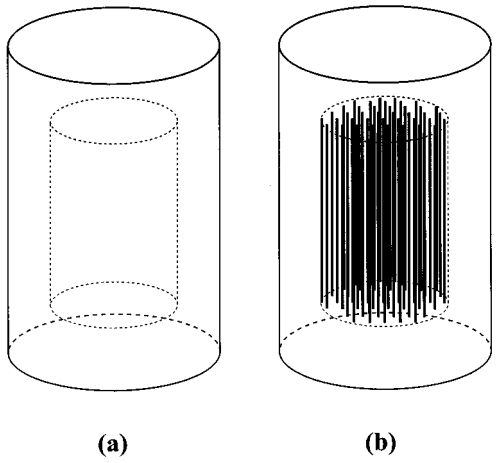


FIG. 1. An illustration of the two vessel configurations used in the development of the theory of the measurement of f_κ. The (a) empty and (b) full containers are identical except for the pin-array stack housed within the latter. The total volume V_{tot} of each has been partitioned in the same manner such that the interior volume V_i indicated by the dashed lines is that which includes the stack in the full configuration. The remaining volume V_e is that which is exterior to the stack.

with $|V_1| \ll V_{\text{tot}}$ so that the acoustic approximation is valid (i.e., so that the pressure response p_1 will be linear in V_1). Furthermore, we assume that the linear dimensions of the containers are small compared to the acoustic wavelength so that the gas within each chamber may be considered to be spatially isobaric at each instant of time.

The nature of the pressure response of the gas will depend upon the ratio of δ_κ to the characteristic dimensions ℓ_i of the vessels. If the oscillations are slow enough that heat can diffuse distances much greater than the linear dimensions of the container during each acoustic period, the gas will respond isothermally. If, on the other hand, δ_κ is much smaller than any of the characteristic dimensions, so that the volume of gas situated within a distance δ_κ of the solid is very small compared to V_{tot}, the gas response will be adiabatic. Between these limits, the response of the gas will be a complex function of ℓ_i/δ_κ. Here, we consider the situation in which δ_κ is much smaller than the characteristic dimensions of the vessel itself, but where it may be comparable to the half-spacing y_0 between adjacent pins in the stack. The pressure response in the empty configuration will thus be primarily adiabatic while the pressure response in the full configuration will vary from adiabatic to isothermal depending on the size of δ_κ relative to y_0. We assume throughout that the thermal inertia of all solids is large enough for the gas–solid interfaces to remain isothermal.

We start with the first-order equation of state for an ideal gas

$$\frac{\rho_1}{\rho_m} = \frac{p_1}{p_m} - \frac{T_1}{T_m}, \tag{4}$$

where ρ_1 is the complex amplitude of the acoustic density oscillations induced by the volume changes and ρ_m is the mean density of the gas. To first order, integration of Eq. (4) over the volumes of the containers leads to

$$\frac{1}{\rho_m}\int_{\text{empty}}\rho_1\,dV = V_{\text{tot}}\frac{p_1}{p_m} - \frac{1}{T_m}\left[\int T_1\,dV_i + \int T_1\,dV_e\right] \quad (5)$$

for the empty configuration and

$$\frac{1}{\rho_m}\int_{\text{full}}\rho_1\,dV = (V_{\text{tot}} - V_{ip})\,\frac{p_1}{p_m} - \frac{1}{T_m}\left[\int T_1\,dV_{ig} + \int T_1\,dV_e\right] \quad (6)$$

for the full configuration. Making use of Eq. (2) as a definition of f_κ in terms of the spatial average of T_1 over the volume V_{ig}, Eqs. (5) and (6) become

$$\frac{1}{\rho_m}\int_{\text{empty}}\rho_1 dV = V_{\text{tot}}\frac{p_1}{p_m} - (V_{ig}+V_{ip})\left(\frac{\gamma-1}{\gamma}\right)\frac{p_1}{p_m} - \frac{1}{T_m}\int T_1\,dV_e \quad (7)$$

and

$$\frac{1}{\rho_m}\int_{\text{full}}\rho_1 dV = (V_{\text{tot}} - V_{ip})\,\frac{p_1}{p_m} - V_{ig}(1-f_\kappa) \times\left(\frac{\gamma-1}{\gamma}\right)\frac{p_1}{p_m} - \frac{1}{T_m}\int T_1\,dV_e, \quad (8)$$

respectively. Note that f_κ does not appear in Eq. (7) because we assume that the gas situated in V_i undergoes adiabatic oscillations in the empty configuration. Dividing Eqs. (7) and (8) by p_1, subtracting one from the other, and recognizing that the first-order equation of continuity implies

$$\int \rho_1 dV = \rho_m V_1 \quad (9)$$

leads to

$$\left(\frac{V_1}{p_1}\right)_{\text{full}} - \left(\frac{V_1}{p_1}\right)_{\text{empty}} = \left(\frac{\gamma-1}{\gamma p_m}\right)\left[f_\kappa V_{ig} - \frac{V_{ip}}{\gamma-1}\right] \quad (10)$$

or

$$f_\kappa = \left(\frac{\gamma}{\gamma-1}\right)\frac{p_m}{V_{ig}}\left[\left(\frac{V_1}{p_1}\right)_{\text{full}} - \left(\frac{V_1}{p_1}\right)_{\text{empty}}\right] + \frac{1}{\gamma-1}\frac{V_{ip}}{V_{ig}}. \quad (11)$$

That is, in principle f_κ can be determined by measuring the pressure response p_1 of the gas to a volume perturbation V_1 for the two configurations.

Successful design of an experiment to measure f_κ based upon the analysis presented above requires careful attention to detail. The key factor V_1/p_1 is a ratio of complex quantities, so phase-sensitive measurements are needed. Unfortunately, the difference between V_1/p_1 for the two geometries is typically small compared to $|V_1/p_1|$. Hence, a very accurate determination of V_1/p_1 is crucial. To ensure the largest difference possible, V_i must occupy most of V_{tot}. The former cannot, however, be chosen at random because the value of f_κ which is inferred from Eq. (11) reflects the spatial average of T_1 over the entire volume V_{ig}. If the pin array is not homogeneous over V_i, then f_κ will be compromised. Such is the case in the illustrative example outlined in Fig. 1 where V_i includes a stack of finite extent. Thermal relaxation processes near the outer periphery and near the ends of the stack will not be the same as those which occur deep within the pin array. In Sec. III we describe two geometries analogous to the empty and full configurations yet carefully chosen such that finite-size effects are moved from V_i to the exterior volume V_e. Finite-size effects thus cancel out when the difference described by Eq. (11) is taken, and the values for f_κ which are obtained correspond to those of a pin array with infinite extent.

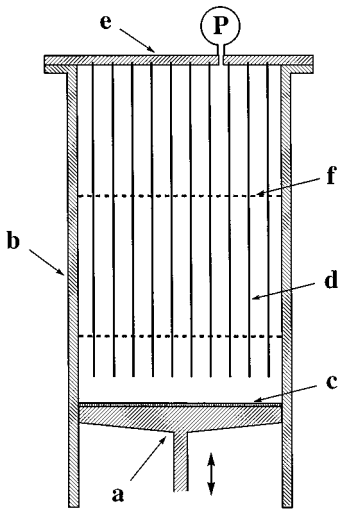


FIG. 2. An outline of the experimental chamber as described in the text. Details include the (a) piston, (b) cylinder, (c) oil seal, (d) pin array, (e) top plate, (f) support plates (2), and (P) pressure gauge. A representative segment of the support plates is illustrated in Fig. 3 while the physical arrangement of pins is shown in Figs. 4 and 5.

II. APPARATUS

A schematic of the apparatus is shown in Fig. 2. Oscillating motion of piston (a) produces changes in the volume of the cylindrical chamber (b). The ensuing pressure variations are detected by the transducer (P). Gas leakage past the piston is prevented by a pool of oil (c) on the upper surface of the piston. The bulk of the chamber volume is occupied by a triangular array of metal rods or ''pins'' (d) individually mounted in and suspended from the upper plate (e). Uniform spacing between the pins is maintained over their length by support plates (f) which are largely open to the flow of gas. A representative segment of one of the support plates is shown in Fig. 3.

The experimental chamber comprises a 394-mm-diam right-circular cylinder approximately 990 mm in length (volume $V_{\text{tot}}=0.12\ \text{m}^3$). It is sealed above by a rigid plate from which the pins are suspended and below by the oil-covered piston. The latter separates the experimental chamber from a second sealed chamber housing a 155-mm-stroke linear motor[21] which drives the piston. The weight of the piston is borne by a small differential pressure between the two chambers. In addition to providing a seal between the two chambers, the oil serves to lubricate the running surfaces between the piston and the cylinder walls. The vapor pressure of the

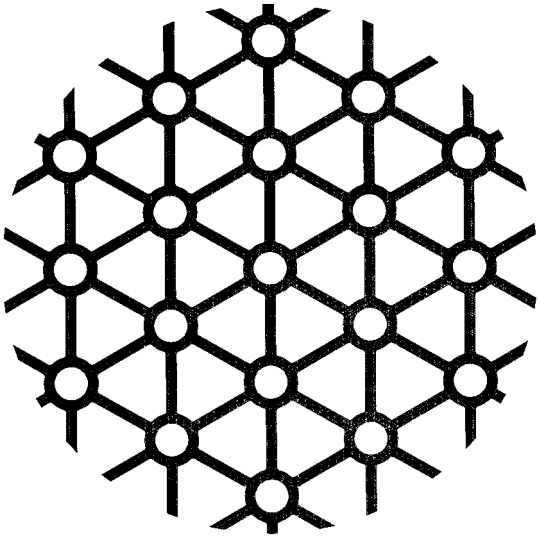


FIG. 3. An illustration of a representative segment of the support plates used to maintain a uniform pin spacing over the length of the pin array. The pins pass through the circular holes while gas flows through the ''triangular'' passages between pins.

oil[22] is low enough to ensure that the properties of the gas are not influenced. A baffle (not shown) located just below the surface of the oil inhibits the generation of surface waves. The walls and the upper plate of the experimental chamber are heavily insulated and were maintained at a constant and uniform temperature of 20.0(1) °C throughout all phases of the experiment.

The mean position of the piston, hence the overall length of the experimental chamber, was measured *in situ* using a calibrated linear variable differential transformer (LVDT) located below the piston. Repeated checks over the course of the experiment indicate that any long-term variability of the absolute calibration of the device was less than 25 μm. The actual length of the chamber increased as a function of time ($\approx$65 μm/day) for a given LVDT reading as oil leaked past the piston. This effect was easily accounted for by periodically measuring the depth of the oil. Typically $V_{\rm tot}$ was reproduced to better than 0.1% accuracy between experimental runs.

The amplitude and phase of the motion of the piston at the acoustic frequency were also detected using the LVDT. Commercially available components were used; however, a number of modifications were made to improve the accuracy and reproducibility of the dynamic response of the system. Small amplitude changes in position are sensitive to the local derivative of the transfer function (output voltage as a function of position) rather than to the mean slope as is the case for large displacements. A layer-wound transformer coil[23] was consequently used to ensure a maximally flat transfer function. The length of this coil was a compromise between the full stroke of the piston and the typical amplitude of the piston motion. The core of the LVDT was rigidly attached to the underside of the piston. A non-magnetic (Ti) rod was used for this purpose to avoid the potential introduction of nonlinearities into the transfer function. An identical Ti rod was attached to the other end of the core to enhance the symmetry of the core support structure. This assembly (coil, core and support rods) was then mounted within two cylindrical magnetic shields to screen out fields generated by the linear motor. The primary of the transformer coil was excited at 10 kHz using a commercially available package.[24] Once assembled, the rigidity of the entire piston-LVDT structure was verified in a separate experiment in which a second, quite different LVDT was mounted to the upper surface of the piston and the responses of both LVDT's to motion of the piston were monitored simultaneously.

Following these modifications, the local sensitivity of the LVDT to small displacements exhibited variations of order 1% with a periodicity of order 1 cm. This residual distortion was accounted for by performing a detailed calibration of the output voltage as a function of position and then calculating a position- and amplitude-dependent correction to the data as they were acquired. Analyzed in this manner, any residual systematic variation in the local sensitivity of the piston position-measurement system was reduced to less than 0.1% over the full operating range of the LVDT.

The oscillating pressure within the experimental chamber was measured using a piezo-resistive transducer[25] mounted in the top plate (and consequently temperature regulated). These readings were verified under a wide range of conditions using auxiliary transducers located at various positions, to assure that the chamber was isobaric as assumed earlier. The back volume of the transducer was connected to a separate, isolated volume maintained at a pressure close to that of the experimental chamber, so that the sensing element was always exposed to the same (small) average stress. The pressure within the latter volume was measured using a temperature-regulated capacitance gauge[26] calibrated against a Hg manometer.[27] The mean pressure within the experimental chamber was thus taken to be the sum of the pressure within the isolated back volume plus the small mean pressure difference across the piezo-resistive transducer. ^{4}He gas at mean fill pressures ranging from 100 to 1200 Torr was used in all phases of the experiment to obtain large thermal penetration depths.

The piston was typically driven at peak amplitudes of order 3 mm, corresponding to fractional volume (or pressure) changes of order 0.3%. Harmonics in the pressure waveform were monitored, and attributed to distortions in the motion of the piston caused by stick-slip processes. Small corrections were made to the linear motor drive current at double and triple the drive frequency so as to eliminate these harmonics in the pressure response, if present. The remaining total harmonic distortion of the LVDT output was monitored continuously and was consistent with nonlinearities in the previously measured transfer function.

Drive frequencies ranging from 2 Hz to 16 Hz were used. The length of the chamber thus varied from 0.2% to 1.6% of an acoustic wavelength. The resulting gas flow velocities were always small enough that the viscous dissipation of power within the chamber could be neglected compared to thermal effects. The maximum Reynolds number based upon the full diameter of the experimental chamber was of order 30 and hence turbulence was not expected to influence the data.

A total of $N_{\mathrm{tot}}=1147$ blind holes were bored into the upper plate of the chamber on a triangular grid with a nearest neighbor center-to-center spacing $2y_0=11.00(2)$ mm (Ref. 28) so as to accommodate the same number of $2r_i=3.18(1)$ mm diameter stainless-steel rods. The latter were affixed using hard wax[29] so that they were easily removed or exchanged. Pins with nominal lengths of 50 mm and 910 mm were used and, as discussed below, a pin was located in each site for all configurations which were considered. A variable number (2 or 3) of spacer grids (Fig. 3) were used to maintain the regularity of the grid pattern over the length of the pins. The maximum measured deviation from a perfect triangular grid was of order 1% of the inter-pin spacing, while the volume-weighted deviation was of order 0.4% of this distance.

The largest uncertainty in the total volume occupied by the pins was simply due to the fact that they varied in length and that the length of each pin was not individually measured. The average pin length was determined to within a standard deviation of 0.8 mm or about 0.1% of the length of the longer pins.

The thermal penetration depth in the pins themselves ranged from about 280 μm to 790 μm under the experimental conditions. The heat capacity of the solid, however, was always much greater than that of the gas near the solid. Consequently the assumption of an isothermal boundary condition at the pin surface was excellent.[30]

III. EXPERIMENTAL PROCEDURE

The oscillating pressure p_1 in the experimental chamber resulting from motion of the piston was measured as a function of frequency and gas pressure; that is, as a function of the thermal penetration depth δ_κ within the gas. This procedure was carried out over a range of δ_κ from about 1.5 mm to 15 mm for a number of pin-stack configurations as described below.

The outputs of the pressure transducer and the LVDT were both measured using a calibrated lock-in amplifier[31] referenced to the fundamental drive frequency of the linear motor. The LVDT signal was accurately attenuated to a level comparable to that of the signal from the pressure transducer so that both could be measured using the same gain setting on the lock-in. Carefully set and monitored bandpass filtering ($Q=50$) was used for both signals, and the reference phase of the lock-in was adjusted to obtain maximum quadrature phase accuracy. The input signals and the quadrature phase switch of the lock-in were then repeatedly toggled under computer control to measure the amplitude and phase relationship between p_1 and V_1. Note that the latter is related to the amplitude of the piston motion as detected by the LVDT via the piston area. The mean pressure and piston location were recorded at the same time.

Measurements under a particular set of conditions were generally carried out over a 24-hour period following an initial equilibration time. The complex normalized ratio of oscillating volume to pressure $(V_1/p_1)(p_m/V_{\mathrm{tot}})$ (Ref. 32) was typically measured to a statistical uncertainty better than 0.1% in amplitude with a relative phase uncertainty less than 0.05 degrees following the few hundred observations which could be obtained during this time. A systematic periodic variation of this quantity with a 24-h period was observed. The amplitude of this effect was of the same order of magnitude as the statistical uncertainty and was attributed to the influence of uncontrolled diurnal temperature changes in the laboratory on electronics external to the experimental chamber. As this effect was highly reproducible, the data were fit to a constant plus an oscillating component which was then subtracted. Data collected in this fashion were reproducible at the 0.1% level over time periods of months and after fully dismantling and reassembling pin configurations.

Investigations were carried out to examine the amplitude dependence of the ratio V_1/p_1. No effects were observed to within the reported experimental uncertainties, confirming the assertions of laminar flow and the validity of the acoustic approximation during the analysis of data. Likewise, harmonics were deliberately introduced into the motion of the piston to allow examination of their potential influence on the interpretation of the data. Sufficient care was taken in the collection of data which are presented here to ensure that they have not been influenced by harmonics at the reported level of uncertainty.

Two basic pin configurations were examined in detail. We refer to these arrangements as ''background'' and ''foreground'' configurations,[33] corresponding roughly to the empty and full configurations of Sec. I. Rather than simply determine the difference in V_1/p_1 for the empty and full configurations of Sec. I, a concerted effort was made to account for and eliminate systematic uncertainties associated with finite-size effects as described below. In effect, geometries were chosen so as to arrive at an unambiguous value for V_{ig}.

Figures 4 and 5 illustrate the full triangular arrangement of pins within the experimental chamber. The large circle which surrounds the pin sites delineates the location of the chamber walls. Despite the large number of pins, roughly 10% of the sites are located within a distance of order $2y_0$ from the walls. Gas situated near these sites does not experience the same thermal relaxation as would be encountered near sites far from the chamber walls. Because of these ''edge'' effects, a simple difference between the full and empty configurations as implied by the analysis of Sec. I is bound to yield a value for f_κ which is different from that of a perfect pin-array stack with infinite extent.

A similar argument can be made with respect to ''end'' effects. That is, one end of the pin array intersects the top plate of the experimental chamber while the other end dangles freely in open space. At neither end of the pin array does the gas experience the same local boundary conditions as does gas situated far from the ends.

The background configuration is illustrated in Fig. 4. Full length ($L=910$ mm) pins were placed at the $N_{\mathrm{back}}=234$ sites indicated with filled symbols, forming a ''guard ring'' around the outer walls of the chamber. The remaining sites (open symbols) were filled with short pins ($l=50$ mm) completing the full triangular pattern near the top plate. Observation of this pattern shows that there are three types of long-pin edge sites around the inner periphery of the guard

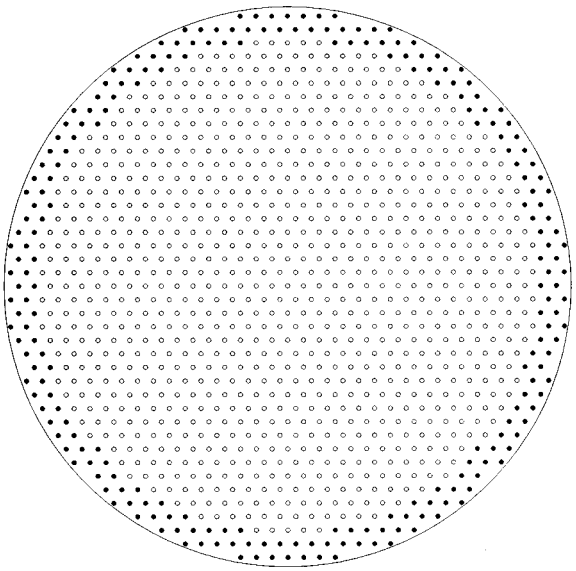


FIG. 4. The triangular arrangement of pins which we refer to as the background configuration. Full length (910 mm) pins are situated at the sites indicated by the filled symbols forming a guard ring around the inner walls of the cylinder (large circle). Short (50 mm) pins are located in the central region at the sites indicated by the open symbols.

ring: sites which have as nearest neighbors, one, two, or three sites occupied by short pins.

The foreground configuration is shown in Fig. 5. Rather than fill all sites with full length pins, a residual void populated by 135 short pins has been retained. Comparison of Figs. 4 and 5 shows that the total number of each of the three types of edge sites has been conserved. Likewise, the total number of pins which intersect the top plate and which hang in free space is the same. As far as first nearest neighbors are concerned, the difference between the background and foreground configurations is due solely to the effective elongation of $N_i=778$ pins located in a perfect triangular configuration of full-length pins. Hence V_{ig} is unambiguously given by

$$V_{ig}=N_i(L-l)[2\sqrt{3}y_0^2-\pi r_i^2], \tag{12}$$

where $2\sqrt{3}y_0^2$ is the area of the hexagonal unit cell around each pin site and πr_i^2 is the cross-sectional area of a pin. The picture breaks down for second nearest neighbors; however, their influence on the data is expected to much less than that of first nearest neighbors.

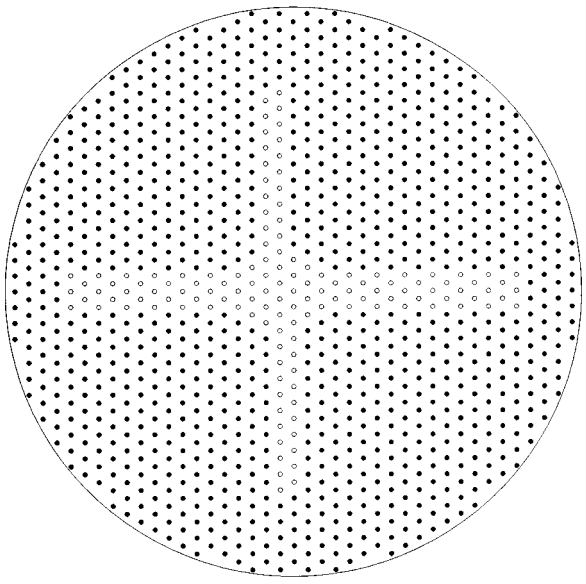


FIG. 5. The triangular arrangement of pins which we refer to as the foreground configuration. As for the background configuration, the placement of the short and full length pins on the triangular grid is indicated by the open and filled symbols, respectively. The central void created by the short pins is designed such that the inner periphery replicates the number and type of sites around the inner periphery of the guard ring of the background configuration.

A series of measurements was first carried out with the pins arranged in the background configuration. Two support plates were installed such that they only intersected the long pins. Changing the location of these plates had no effect on the measured value of p_1 for a given V_1 to within the experimental uncertainty as expected. For future reference, we label this configuration ''background *a*.''

The arrangement of the pins was then changed to the foreground configuration, again with two support plates in place, and p_1 was re-measured under the identical set of conditions for which the background configuration was examined. Typically p_m was reproduced at the level of a few tenths of a percent between configurations. Once again, the influence of the location of the support plates was examined and determined to be insignificant. The pin configuration was then returned to the original background array to verify that no systematic changes had occurred over the course of the experiment.

Unfortunately, direct comparison of the data from the foreground configuration with that from background *a* is obscured by the need for support plates to maintain uniform pin spacing. In particular, the assignation of a well defined value to the volume V_{ig} becomes difficult with support plates in place. Recall that V_{ig} is that volume of gas which experiences thermal relaxation as described by f_κ for a pin stack in the full configuration, and which undergoes adiabatic temperature oscillations in the empty configuration. Clearly, without the support plates V_{ig} is given by Eq. (12). With the addition of the support plates V_{ig} is reduced by a volume of order $2\delta_\kappa V_{ig}/(L-l)$ per plate. That is, gas situated within a distance of order one thermal penetration depth either side of the support plates experiences non-adiabatic temperature oscillations in *both* configurations. Unfortunately the nature of the non-adiabaticity is different in the two cases and a simple correction to account for the change in V_{ig} with δ_κ is not adequate.[34] In the present work we account for the uncertainty in V_{ig} introduced by the support plates in a perturbation fashion. That is, by changing the number and the placement of support plates as outlined below, we are able to account for and eliminate their influence on the measured pressure response of the gas.[35]

Following the basic set of measurements described earlier, the background configuration was modified such that one of the support plates intersected *all* of the pins rather than just the long pins. We label this configuration ''back-

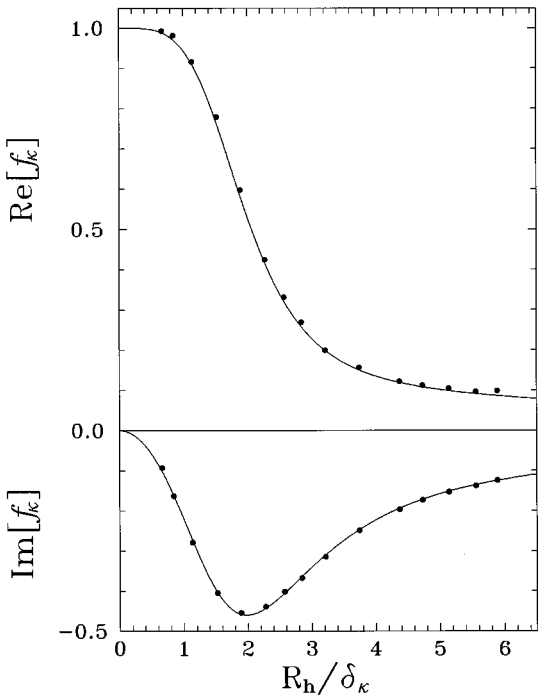


FIG. 6. The real and imaginary components of the function f_κ as measured in this experiment. The solid curves correspond to the theoretical prediction of Swift and Keolian (Ref. 6). The data have been plotted against the ratio R_h/δ_κ for ease of comparison with other stack geometries.

ground *b*'' to distinguish it from background *a*. The pressure response of the gas to volume perturbations was once again measured under conditions identical to those used previously. A small but experimentally significant difference in V_1/p_1 relative to the data obtained with background *a* was observed. This difference was attributed to the difference in the nature of the non-adiabaticity of the acoustic oscillations near the support plates for the two cases. That is, in one case $N_{\text{back}}=234$ pins (i.e., the guard ring) intersect the support plate while in the other all $N_{\text{tot}}=1147$ pins pass through it. We assume that the proper background pressure response to which the foreground pressure response should be compared is obtained by scaling the difference between the results for the two backgrounds according to the total number of pins that intersect the support plates:[36]

$$\left(\frac{V_1}{p_1}\right)_{\text{bkgd}} = \left(\frac{V_1}{p_1}\right)_{\text{bkgd } a} + \frac{2N_i}{N_{\text{tot}}-N_{\text{back}}} \times\left[\left(\frac{V_1}{p_1}\right)_{\text{bkgd } b} - \left(\frac{V_1}{p_1}\right)_{\text{bkgd } a}\right]. \tag{13}$$

Limited data were also collected in the foreground configuration and both background configurations with three support plates in place. Analysis of these data in an analogous manner yields identical results for f_κ, allaying any concerns that differences in the pin distribution homogeneity contributed to the difference in the pressure response between background configurations.

Our results for f_κ are plotted in Fig. 6 as a function of the ratio R_h/δ_κ where R_h is the hydraulic radius[37]

$$R_h=\frac{2\sqrt{3}y_0^2-\pi r_i^2}{2\pi r_i} \tag{14}$$

which for our geometry is $R_h=9.69$ mm. The data points correspond to values of f_κ calculated from Eq. (11) interpreting the foreground configuration as the full configuration of Sec. I and the background given by Eq. (13) as the corresponding empty configuration. The included gas volume V_{ig} is taken to be exactly the difference in pin-stack volumes between the two configurations as implied by Eq. (12). Results are presented only for the case where two support plates were used. Results obtained with three plates cannot be distinguished from those shown in the figure. Values for δ_κ were calculated using $\kappa=(1.76\ \text{cm}^2/\text{s bar})\cdot p_m$ for the thermal diffusivity of the gas while values calculated for f_κ assume $\gamma=5/3$. The absolute uncertainty in each datum is of order 0.01 for both components of f_κ. One should note that the perturbing influence of the support plates was indeed greatest for small R_h/δ_κ as expected. Neglect of this effect would have led to an error of order 0.04 in $\text{Im}[f_\kappa]$ and 0.10 in $\text{Re}[f_\kappa]$ for the smallest values of R_h/δ_κ.

IV. SUMMARY AND CONCLUSIONS

We have taken advantage of the principles of similitude to build and study the performance of a large-scale model of a thermoacoustic pin-array stack. In this manner we were able to ensure a high degree of geometrical uniformity over the entire stack volume. Measurements of the steady-state pressure response p_1 of a fixed quantity of gas confined to a sealed vessel were made as the volume of the chamber was subjected to sinusoidal perturbations. The response of the gas was modified when the stack was placed within the chamber. This change was interpreted in terms of the function f_κ which relates the spatial average of the amplitude of the oscillating component of temperature to that of pressure. This function is of considerable practical interest in that it characterizes the performance of a thermoacoustic stack.

The results of our measurements are summarized in Fig. 6, in which f_κ is plotted as a function of the ratio of the hydraulic radius R_h defined by Eq. (14) to the thermal penetration depth δ_κ. Also plotted in this figure is the theoretical prediction of Ref. 6 for f_κ. The agreement between theory and experiment is excellent.

As a result of this agreement the data can be interpreted as a practical demonstration of the inherent superiority of the pin-array geometry over other thermoacoustic stack designs. Swift and Keolian introduced a tentative figure of merit

$$M=\sqrt{\sigma}\ \text{Im}[f_\kappa]\ \frac{|1-f_\nu|^2}{\text{Im}[f_\nu]} \tag{15}$$

for comparing different stack geometries.[6] Here, σ is the Prandtl number, and f_ν is given by f_κ after substituting $\delta_\nu=\sqrt{\sigma}\delta_\kappa$ for δ_κ.[38] Their choice of parameters was based on the observations of Arnott *et al.*[12] that thermoacoustic heat transport and work are proportional to $\text{Im}[f_\kappa]$ in the inviscid, standing wave limit, while the acoustic dissipation of power by viscosity is proportional to $\text{Im}[f_\nu]/|1-f_\nu|^2$ in the absence of thermal gradients. From a practical standpoint,

one is interested in comparing the value of M exhibited by various geometries at that value of R_h/δ_κ for which thermoacoustic processes are maximized. We denote this alternate figure of merit $\tilde{M}$ where $\tilde{M}$ is the value of M when $|\mathrm{Im}[f_\kappa]|$ is maximum. Based upon this admittedly simplistic criterion, the particular stack studied in this experiment is 23% more efficient than the optimum parallel-plate stack operating at the same gas pressure and frequency, and 51% more efficient than the optimum circular-pore-array stack.

It is interesting to note that the agreement between theory and experiment displayed in Fig. 6 is comparable over the full range of R_h/δ_κ which was examined. In the boundary-layer limit ($R_h/\delta_\kappa \to \infty$) the boundary condition between adjacent hexagonal cells should not influence temperature oscillations near the pin. *A priori* one expects the two to agree in this limit. The fact that agreement is also obtained when $R_h/\delta_\kappa \lesssim 1$ is less intuitive. That is, one expects the boundary conditions between unit cells to have a more pronounced influence on acoustic temperature variations when δ_κ is large. For our measurement at $R_h/\delta_\kappa = 0.66$, the unit cell boundary is only $\delta_\kappa/4$ away from the surface of the pin. In effect, our data make an important contribution to the understanding of pin-array stacks as they extend into a region of parameter space where there is little rigorous justification for the key assumption behind the only existing theory.

In scaling the geometry of the model and subsequently interpreting the pressure response of the enclosed gas volume, we were forced to deal with effects associated with the finite extent of the stack. The principle innovation introduced here is the manner in which these effects were treated. That is, two pin-array configurations were chosen such that the difference in the pressure response of the gas between the two cases was unambiguously related to the function f_κ for a pin array of infinite extent. Extension of these ideas to a wide range of geometries is straightforward and should allow one to examine situations for which theoretical calculations are difficult. One might consider, for example, pin-array stacks fabricated using pins with square cross sections and/or constructed on a square lattice. The same techniques could also be extended to studies in which motion of the gas and the viscous dissipation of power are important, including situations where the acoustic wave vector is perpendicular to the pin-array axis. Such extensions might require two pistons (one at either end of the cylinder), a moving stack, or some combination thereof, to achieve large amplitude motion of the gas relative to the stack.

ACKNOWLEDGMENTS

The authors gratefully acknowledge conversations with Larry Wilen which led to the inception of this experiment. This work was supported by the Office of Fossil Energy in the U.S. Department of Energy, through the Morgantown Energy Technology Center.

[1] G. W. Swift, ''Thermoacoustic engines,'' J. Acoust. Soc. Am. **84**, 1145–1180 (1988).
[2] G. W. Swift, ''Thermoacoustic engines and refrigerators,'' Phys. Today **48**, 22–28 (July 1995).
[3] The stack is similar to a regenerator in that it acts as a thermal reservoir which can interact with the gas. The essential difference between the two lies in the degree to which the gas is coupled to the reservoir. In a stack, the gas is in *imperfect* thermal contact with the solid whereas in a regenerator the typical pore dimension is much smaller and the coupling between gas and solid is consequently much stronger.
[4] T. J. Hofler, Concepts for thermoacoustic refrigeration and a practical device, Proceedings of the 5th International Cryocoolers Conference, Monterey CA (Wright-Patterson AFB, Wright-Patterson, OH, 1998), pp. 93–101; T. J. Hofler, ''Thermoacoustic refrigerator design and performance,'' Ph.D. thesis, Univ. of California, San Diego, 1986.
[5] G. W. Swift, ''Analysis and performance of a large thermoacoustic engine,'' J. Acoust. Soc. Am. **92**, 1551–1563 (1992).
[6] G. W. Swift and R. M. Keolian, ''Thermoacoustics in pin-array stacks,'' J. Acoust. Soc. Am. **94**, 941–943 (1993).
[7] U. A. Müller, ''Thermoacoustic device,'' U.S. Patent number 4,625,517 (1986).
[8] Useful thermoacoustic processes occur at distances of order δ_κ away from the solid while viscous shear is most intense at the gas–solid interface.
[9] R. J. Gibson, F. S. Nessler, and R. M. Keolian, ''Measurements of a thermoacoustic pin stack (abstract),'' J. Acoust. Soc. Am. **100**, 2845 (1996); ''A thermoacoustic pin stack for improved efficiency,'' to be published in the proceedings of the National Heat Transfer Conference, Baltimore MD, August 1997.
[10] G. Kirchhoff, ''Ueber den Einfluss der Wärmeleitung in einem Gase auf die Schallbewegung,'' Ann. Phys. Chem. **134**, 177–193 (1868).
[11] N. Rott and G. Zouzoulas, ''Thermally driven acoustic oscillations. Part IV. Tubes with variable cross section,'' Z. Angew. Math. Phys. **27**, 197–224 (1976).
[12] W. Pat Arnott, Henry Bass, and Richard Raspet, ''General formulation of thermoacoustics for stacks having arbitrarily shaped pore cross sections,'' J. Acoust. Soc. Am. **90**, 3228–3237 (1991).
[13] F. Mechel, ''Eine Modelltheorie zum Faserabsorber Teil I: Reguläre Faserordnung,'' Acustica **36**, 53–64 (1976/77).
[14] F. Mechel, ''Eine Modelltheorie zum Faserabsorber Teil II: Absorbermodell aus Elementarzellen und numerische Ergebnisse,'' Acustica **36**, 65–89 (1976/77).
[15] J. R. Olson and G. W. Swift, ''Similitude in thermoacoustics,'' J. Acoust. Soc. Am. **95**, 1405–1412 (1994).
[16] N. Rott, ''Damped and thermally driven acoustic oscillations in wide and narrow tubes,'' Z. Angew. Math. Phys. **20**, 230–243 (1969).
[17] L. A. Wilen, ''Experimental measurements of scaling properties for acoustic propagation in a single pore,'' J. Acoust. Soc. Am. **101**, 1388–1397 (1997); L. A. Wilen, ''Measurements of thermoacoustic functions for single pores,'' submitted to J. Acoust. Soc. Am.
[18] H. S. Roh, W. P. Arnott, J. M. Sabatier, and R. Raspet, ''Measurement and calculation of acoustic propagation constants in arrays of small air-filled rectangular tubes,'' J. Acoust. Soc. Am. **89**, 2617–2624 (1991).
[19] M. R. Stinson and Y. Champoux, ''Propagation of sound and the assignment of shape factors in model porous materials having simple pore geometries,'' J. Acoust. Soc. Am. **91**, 685–695 (1992).
[20] Y. Champoux and M. R. Stinson, ''Measurement of the characteristic impedance and propagation constant of materials having high flow resistivity,'' J. Acoust. Soc. Am. **90**, 2182–2191 (1991).
[21] APS Dynamics (Carlsbad, CA 92000) model 400S.
[22] Inland Vacuum, Inc. (Churchville, NY 14428) type 19: 2×10^{-5} Torr at 20 °C.
[23] Lucas Control Systems Products (Hampton, VA 23666) model M12-50.
[24] Lucas Control Systems Products (Hampton, VA 23666) model ATA-101: internal filters were modified to a 3-pole Bessel configuration with a corner frequency of 1 kHz.
[25] Endevco Corp. (San Juan Capistrano, CA 92675) model 8510B-5.
[26] MKS Instruments, Inc. (Andover, MA 01810) model 390HA-1000 Baratron gauge.
[27] The calibration of each piezo-resistive transducer was likewise checked against the manometer.
[28] The cumulative uncertainty in the absolute distance between any pair of holes in both the top plate and the support plates was maintained at the same level.
[29] Argueso, M., & Co., Inc. (Mamaroneck, NY 10543) type Rigidax WS.
[30] The effect of non-isothermal boundary conditions is described (Refs. 6, 1) in terms of the complex dimensionless parameter ϵ_s. Throughout the present experiment, $|\epsilon_s|$ was never greater than $4\times10^{-3} \ll 1$.

[31] EG&G Instruments Corp. (Princeton, NJ 08543) model PAR 124A.
[32] The magnitude of this quantity is of order unity.
[33] The nomenclature used to describe these carefully contrived pin arrangements is intended to distinguish them from the abstract ''empty'' and ''full'' configurations of Sec. I, and yet give some indication of the implied analogies.
[34] The magnitude of this correction is small when $\delta_\kappa \ll y_0$ and the total surface area of the support plates is much less than that of the pins.
[35] This approach relies upon the relative change in $(V_1/p_1)(p_m/V_{\rm tot})$ being small as individual support plates are added or removed from the chamber. Each plate causes V_{ig} to be decreased by a factor of order $2\delta_\kappa/(L-l)$ which, for δ_κ of order the inter-pin spacing, is only a few percent in the present experiment.
[36] Recall that only $N_{\rm fore}=1012$ out of $N_{\rm tot}=1147$ pins intersect the support plates in the foreground configuration. Equation (13) is specific to the case of two support plates.
[37] The hydraulic radius is the ratio of the cross-sectional area of the gas to the perimeter of the solid. The data have been plotted in this fashion for ease of comparison between stack geometries.
[38] The viscous penetration depth δ_ν is analogous to δ_κ in that it is the characteristic length scale for the diffusion of momentum at the acoustic frequency.

A model of ultrasonic nondestructive testing for internal and subsurface cracks

Peter Bövik
University of Trollhättan/Uddevalla, Box 957, 461 29 Trollhättan, Sweden

Anders Boström
Division of Mechanics, Chalmers University of Technology, S-412 96 Göteborg, Sweden

(Received 27 February 1997; accepted for publication 15 July 1997)

The scattering of elastic waves in a half-space containing a striplike crack is investigated. As a special case it seems that the crack may be surface breaking. A surface integral equation with the half-space Green tensor is employed. The key point of the method is the expansion of the Green tensor in Fourier representations with the free part of the Green tensor expanded in the crack coordinate system and the half-space part in the half-space coordinate system. The integral equation is discretized by expanding the crack opening displacement in terms of Chebyshew functions having the correct square root behavior along the crack edges. The incident field is emitted from an ultrasonic probe and a recent model for this is employed. The signal response in another (or the same) probe is modeled by a reciprocity argument and the stationary phase approximation is employed to simplify the final answer, which is thus only valid in the far field of the probes (yielding essentially a spherical wave). Numerical results are given and are compared with both other methods and with available experiments. © *1997 Acoustical Society of America.*
[S0001-4966(97)01011-4]

PACS numbers: 43.35.Zc, 43.35.Cg, 43.20.Gp [HEB]

INTRODUCTION

Ultrasonic nondestructive testing (NDT) of various types of materials and components is routinely used in many branches of industry, examples being the aerospace and nuclear power industries. During the last two decades the development of mathematical models of the testing situation has led to many more or less refined models of various particular cases. There are many benefits from an accurate and validated model. A good model is useful when developing testing procedures as it is easy and cheap to perform parametric studies. For educational purposes and for the enhancement of physical understanding and intuition a model is almost indispensable. As a model can generate noise-free signals, its output is useful when developing signal processing and data inversion procedures.

The most important type of defect to find during testing is usually various types of cracks as these may lead to catastrophic failure under unfavourable conditions. There now exist some models that treat the scattering by cracks and include more or less accurate models of the ultrasonic probes. Thus Chapman[1,2] employs the geometrical theory of diffraction (GTD) to treat some types of internal cracks. Purely numerical methods where a volume discretization is performed are also used more and more. Examples include the work by Fellinger *et al.*,[3] Lord *et al.*,[4] Harumi and Uchida,[5] Nouailhas *et al.*,[6] and Minachi *et al.*[7] These methods can be used for almost any geometry including anisotropic and inhomogenous materials, but have some serious limitations. The number of nodes, etc., easily becomes excessively large, so typically these methods are presently only used in two dimensions (including most of the abovementioned references). In three dimensions ultrasonic NDT problems typically lead to 10^6-10^9 unknowns, and usually this is too many unknowns for reasonable execution times and memory requirements on present computers. Approximate theories that models the ultrasonic NDT situation have been proposed by Lawn *et al.*[8] and Lhémery,[9] among others.

In the present paper a recently developed model[10] of ultrasonic NDT for some simple internal defects (spheres, spheroids, and circular cracks) is extended to include a striplike crack that may also be located close to a free back surface or may even approach the surface and essentially be surface breaking. This geometry is of particular interest in the nuclear power industry where thick-walled components like pipes are quite common. The model does not employ volume discretization or approximate solutions; instead integral equation methods are used to solve the scattering problems. This only works for special defects, but it has the virtue of solving the defect scattering exactly with reasonable computer efforts. It is also possible to treat quite large defects with dimensions above ten wavelengths.

Surface breaking and subsurface cracks are considered by a number of authors,[4–7,11–18] where further references may also be found. Also distributions of surface breaking cracks have been considered; see, e.g., Zhang and Achenbach.[19] The methods employed are either of the volume discretization type or of integral equation type, mostly boundary element methods. As opposed to most of the approaches, the method used in the present paper treats the three-dimensional problem. The present integral equation is also novel in that it uses the half-space Green tensor which is Fourier transform represented. The key point is that the free part of the Green tensor is represented in the crack coordi-

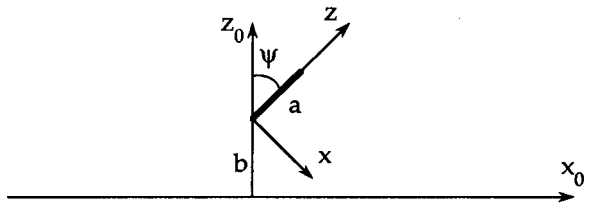


FIG. 1. The geometry of the crack and the free surface.

nate system, while the half-space part is represented in the half-space coordinate system. The method for treating the free part is in fact the same as the one used by Guan and Norris[20] to investigate the rectangular crack in a full elastic space. This method is in turn equivalent to the method of Krenk and Schmidt[21] for solving the scattering by a circular crack, see also Boström[22] for the acoustic scattering by a sound-hard rectangle.

The plan of the present paper is as follows. In the next section the scattering problem for the striplike crack in a half-space is first stated. The integral equation is derived and the Fourier representation of the Green tensor is discussed in some detail. The integral equation is discretized by expanding the unknown crack opening displacement (COD) in terms of Chebyshew functions having the correct square root behavior along the crack edges. In Sec. II the incoming probe field and its transformation to the crack coordinate system are discussed. In Sec. III the reciprocity argument that is used to model the signal response in the receiving probe is given and the stationary phase approximation is employed to get rid of the double integrals that appear in the signal response. This approximation is only valid in the probe far field, which is essentially a spherical wave. Numerical results and comparisons with experiments and other models are considered in Sec. IV. Comparisons are performed both for internal cracks without back surface and for surface breaking cracks. Generally speaking the agreement is quite good. Inaccuracies that may result from the stationary phase approximation are discussed.

I. THE INTEGRAL EQUATION

Consider a scattering geometry with a striplike crack in the vicinity of a free planar surface, cf. Fig. 1. The tilt of the crack relative the surface normal is denoted by ψ ($\in[-\pi/2,\pi/2]$) and the distance from the surface to the closest crack edge is b. As a special case the crack may be surface breaking, i.e., $b=0$. The width of the crack is a. The crack coordinate system xyz has the origin at the closest crack edge, the x axis is normal to the crack, the y axis is along the crack edge, and the z axis is pointing across the crack. The free surface coordinate system $x_0y_0z_0$ has the origin on the free surface at the point closest to the xyz origin. The z_0 axis is normal to the free surface and the y_0 axis is parallel to the y axis.

The material in the half-space is assumed to be homogenous and isotropic, linearly elastic with Lamé constants λ and μ and density ρ. Only time-harmonic conditions are treated and the factor $\exp(-i\omega t)$, where ω is the angular frequency, is suppressed throughout. The longitudinal and transverse wave numbers are k_p and k_s, respectively.

The boundary condition on the free surface of the half-space is of course that the traction vector vanishes there. On the crack the more general spring boundary conditions are assumed. Thus the traction is continuous across the crack and proportional to the crack opening displacement (COD):

$$\mathbf{t}_+ = \mathbf{t}_- = \alpha \Delta \mathbf{u}. \tag{1}$$

Here $\mathbf{t}$ denotes the traction, $\Delta\mathbf{u}$ the COD, and index $+$ or $-$ the limiting values from the positive or negative side, respectively. The quantity α is allowed to be a tensor with components $\alpha_{nn'}$. In practice there seems to be little reason to have a nondiagonal α. By proper choice of α the crack may be open, fluid filled, partly closed, or a thin layer of a softer material. For an open crack $\alpha=0$.

The incoming field is not specified in this section but left arbitrary. In the next section the incoming field is further discussed. It should be noted that it is the incoming field that makes the scattering problem three dimensional, as the geometry can be said to be two dimensional (it is translation invariant in the y direction). To get a simpler integral equation the incoming field is chosen to satisfy the traction-free boundary condition on the surface of the half-space. The incoming displacement field is denoted $\mathbf{u}^i$ and the scattered field is $\mathbf{u}^s$. The total field is then $\mathbf{u}=\mathbf{u}^i+\mathbf{u}^s$.

To derive the integral equation, the starting point is the integral representation for the total displacement field:

$$\mathbf{u}^i(\mathbf{r}) + \frac{k_s}{\mu}\int_{-\infty}^{\infty} dy' \int_0^a dz'\ \Delta\mathbf{u}(\mathbf{r}')\cdot\mathbf{t}'(\mathbf{G}(\mathbf{r}',\mathbf{r})) = \begin{cases} \mathbf{u}(\mathbf{r}), & z_0>0, \\ 0, & z_0<0. \end{cases} \tag{2}$$

Here, $\mathbf{G}$ is the half-space Green tensor, $\mathbf{t}'$ is the traction operator acting on the integration variables, and the dot denotes scalar multiplication. As both the displacement field and the Green tensor satisfy the traction-free boundary condition on the surface of the half-space, there is no surface integral over the boundary of the half-space. The Green tensor satisfy the following equation

$$k_p^{-2}\nabla\nabla\cdot G(\mathbf{r}^-,\mathbf{r}') - k_s^{-2}\nabla\times(\nabla\times G(\mathbf{r},\mathbf{r}')) + G(\mathbf{r},\mathbf{r}') = k_s^{-3} I\,\delta(\mathbf{r}-\mathbf{r}'). \tag{3}$$

Here, I is the unit tensor and the factor k_s^{-3} has been inserted to make G dimensionless.

Let the field point $\mathbf{r}$ be inside the half-space, operate with the traction operator $\mathbf{t}$ on Eq. (2), and let the field point go to the crack surface to obtain

$$\lim_{x\to 0}\int_{-\infty}^{\infty} dy' \int_0^a dz'\ \Delta\mathbf{u}(\mathbf{r}')\cdot\Sigma(\mathbf{r}',\mathbf{r}) = \mathbf{t}_-^s = \alpha\Delta\mathbf{u} - \mathbf{t}^i. \tag{4}$$

In the last step the boundary condition, Eq. (1), is used. The traction Green tensor Σ is obtained by operating with traction operators for both variables on the Green tensor. For definiteness the limit has been taken from $x<0$, but the same

result is obtained if $x>0$ is chosen instead. The limit cannot be taken inside the integrations as the integral equation is of the hypersingular type, cf. Martin and Rizzo.[23] The way the integral equation is solved here in a natural way leads to regular integrals so that the limit can at a later stage be moved inside the integrals.

The integral equation is solved by expanding the Green tensor in plane waves (another way to put this is to say that the Green tensor is expressed as a double Fourier transform) and the COD in Chebyshew functions. The key trick is then to realize that the free (singular) part of the Green tensor should be expanded in plane waves in the xyz system (the crack coordinate system) and the half-space part in the $x_0y_0z_0$ system. The expansion is

$$G_{n'n}(\mathbf{r}',\mathbf{r}) = 2i\sum_j \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{dq\ dp}{k_j h_j} f_{jn'} f_{jn}^* e^{i(h_j|x'-x|+p(y'-y)+q(z'-z))} + 2i\sum_{jj'} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{dq\ dp}{k_j h_j} g_{jn'} R_{jj'} g_{j'n}^+ e^{i(q(x_0'-x_0)+p(y_0'-y_0)+h_j z_0'+h_{j'} z_0)}. \quad (5)$$

Here n and n' run through 1, 2, 3, and j and j' are summed over 1, 2, 3. The wave number $k_j = k_s$ for $j=1,2$ and $k_3 = k_p$. The normal wave number $h_j = \sqrt{k_j^2 - s^2}$ with $s^2 = q^2 + p^2$ and the branch chosen so that $\mathrm{Im}\, h_j \geqslant 0$. For a fixed j, f_{jn} is a vector $\mathbf{f}_j$ with components in the xyz system:

$$\mathbf{f}_1 = \frac{i}{4\pi s}(0,q,-p), \quad (6a)$$

$$\mathbf{f}_2 = \frac{1}{4\pi k_s s}(s^2,-ph_s,-qh_s), \quad (6b)$$

$$\mathbf{f}_3 = \frac{i}{4\pi}\sqrt{\frac{k_p}{k_s^3}}(h_p,p,q). \quad (6c)$$

$\mathbf{f}_j^*$ is obtained by changing ''i'' to ''$-i$'' in $\mathbf{f}_1$ and $\mathbf{f}_3$. Similarly, g_{jn} is a vector $\mathbf{g}_j$ with components in the $x_0y_0z_0$ system

$$\mathbf{g}_1 = \frac{i}{4\pi s}(p,-q,0), \quad (7a)$$

$$\mathbf{g}_2 = \frac{1}{4\pi k_s s}(-qh_s,-ph_s,s^2), \quad (7b)$$

$$\mathbf{g}_3 = \frac{i}{4\pi}\sqrt{\frac{k_p}{k_s^3}}(q,p,h_p). \quad (7c)$$

$\mathbf{g}_j^+$ is obtained by changing all ''h_j'' to ''$-h_j$'' and ''i'' to ''$-i$.'' The reflexion matrix for the free surface has components

$$R_{11} = 1, \quad (8a)$$

$$R_{22} = R_{33} = (4s^2 h_s h_p - K^2)/R, \quad (8b)$$

$$R_{23} = 4i\sqrt{\frac{k_s}{k_p}}\, s h_s K/R, \quad (8c)$$

$$R_{32} = 4i\sqrt{\frac{k_p}{k_s}}\, s h_p K/R, \quad (8d)$$

$$R_{12} = R_{21} = R_{13} = R_{31} = 0. \quad (8e)$$

Here R is the Rayleigh function

$$R = 4s^2 h_s h_p + K^2 \quad (9a)$$

and

$$K = 2s^2 - k_s^2. \quad (9b)$$

It should be stressed that the Fourier transforms in the two terms in Eq. (5) both are with respect to the tangential coordinates for the crack and the surface of the half-space, respectively. For the second term (the half-space term) this is the only possibility, and for the first term it is essential because it means that the awkward term $|x'-x|$ in the exponential vanishes when both x and x' are on the crack i.e., $x = x' = 0$.

The traction Green tensor is obtained from Eq. (5) by applying traction operators with respect to both variables. As the traction Green tensor enters in a surface integral over the crack it is convenient to transform the second term to the crack system xyz at the same time. The result is

$$\Sigma_{n'n}(\mathbf{r}',\mathbf{r}) = 2i\mu^2\sum_j \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{dq\ dp}{k_j h_j} F_{jn'} F_{jn}^* e^{i(h_j|x-x'|+p(y'-y)+q(z'-z))} + 2i\mu^2\sum_{jj'} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{dq\ dp}{k_j h_j} G_{jn'} R_{jj'} G_{j'n}^+ \exp[i(x'(q\ \cos\ \psi + h_j\ \sin\ \psi) - x(q\ \cos\ \psi + h_{j'}\ \sin\ \psi) + p(y'-y) + z'(h_j\ \cos\ \psi + q\ \sin\ \psi) + z(h_{j'}\ \cos\ \psi - q\ \sin\ \psi) + b(h_j + h_{j'}))]. \quad (10)$$

For a fixed j, F_{jn} and G_{jn} are both vectors with components in the xyz system:

$$\mathbf{F}_1 = \frac{1}{4\pi s}(0,-qh_s,ph_s), \quad (11a)$$

$$\mathbf{F}_2 = \frac{i}{4\pi k_s s}(2s^2 h_s,pK,qK), \quad (11b)$$

$$\mathbf{F}_3 = \frac{1}{4\pi}\sqrt{\frac{k_p}{k_s^3}}(K,-2ph_p,-2qh_p), \quad (11c)$$

with the change ''i'' to ''$-i$'' in $\mathbf{F}_j^*$, and

$$G_{11}=\frac{p}{4\pi s}\,(h_s \sin 2\psi-2q\cos^2\psi), \tag{12a}$$

$$G_{12}=\frac{1}{4\pi s}\,[(q^2-p^2)\cos\psi-qh_s\sin\psi], \tag{12b}$$

$$G_{13}=-\frac{p}{4\pi s}\,(h_s\cos 2\psi+q\sin 2\psi), \tag{12c}$$

$$G_{21}=\frac{i}{4\pi k_s s}\,(2s^2h_s\sin^2\psi-2q^2h_s\cos^2\psi - qK\sin 2\psi), \tag{12d}$$

$$G_{22}=-\frac{ip}{4\pi k_s s}\,(2qh_s\cos\psi-K\sin\psi), \tag{12e}$$

$$G_{23}=\frac{i}{4\pi k_s s}\,(qK\cos 2\psi-h_s(s^2+q^2)\sin 2\psi), \tag{12f}$$

$$G_{31}=\frac{1}{4\pi}\sqrt{\frac{k_p}{k_s^3}}\,(2k_p^2\cos^2\psi-k_s^2-2q^2\cos 2\psi + 2qh_p\sin 2\psi), \tag{12g}$$

$$G_{32}=\frac{p}{2\pi}\sqrt{\frac{k_p}{k_s^3}}\,(h_p\sin\psi-q\cos\psi), \tag{12h}$$

$$G_{33}=\frac{1}{4\pi}\sqrt{\frac{k_p}{k_s^3}}\,[(h_p^2-q^2)\sin 2\psi-2qh_p\cos 2\psi], \tag{12i}$$

with the changes ''i'' to ''$-i$'' and ''h_j'' to ''$-h_j$'' in $\mathbf{G}_j^+$.

For large q, the components of $\mathbf{F}_j$ and $\mathbf{G}_j$ behave as $O(q^2)$ and the elements R_{ij} $(i,j=2,3)$ also behave as $O(q^2)$. But because of cancellations in the j and j' sums the asymptotic behaviour of the integrands is (note that for large q the exponential factors are j independent to lowest order as $h_j\simeq i|q|$)

$$\sum_j \frac{1}{k_j h_j}\,F_{jn}F_{jn'}^*=O(q), \tag{13a}$$

$$\sum_{jj'} \frac{1}{k_j h_j}\,G_{jn'}R_{jj'}G_{j'n}^+=O(q). \tag{13b}$$

This behavior is soon needed when the convergence properties of the integrals in the Q matrix are discussed.

For the COD an expansion in Chebyshew functions in the z coordinate and a Fourier transform in the y coordinate is a natural choice:

$$\Delta u_{n'}(y',z')=\sum_{m'}\int_{-\infty}^{\infty}\frac{dp'}{k_s}\,\beta_{n'm'}(p')\phi_{m'}(z')e^{-ip'y'}, \tag{14}$$

where m' (and later also m) is summed from 1 to ∞, $\beta_{n'm'}(p')$ are unknown expansion functions to be determined and

$$\phi_m(z)=\begin{cases} -\dfrac{1}{m\pi}\cos(m\,\arcsin(2z/a-1)), & m=1,3,..., \\ \dfrac{i}{m\pi}\sin(m\,\arcsin(2z/a-1)), & m=2,4,... . \end{cases} \tag{15}$$

This complete set of Chebyshew functions is convenient in the present problem because the functions have the correct square root behavior at the crack edges and the following integral can be computed analytically

$$\int_0^a \phi_m(z)e^{i\gamma z}\,dz=\frac{i}{\gamma}\,e^{i\gamma a/2}J_m(\gamma a/2), \tag{16}$$

where γ is a (possibly complex) parameter. If the crack is allowed to approach the free surface and become surface breaking the behavior at one of the edges changes. It seems that it is still possible to use the same expansion, however, at least in some cases. This is further discussed in Sec. IV.

Inserting the expansion (15) into the integral equation (4), taking a Fourier transform with respect to y, projecting on the set of Chebyshew functions, and using Eq. (16) and the completeness of the Fourier transform leads to

$$\sum_{n'm'} Q_{nmn'm'}(p)\beta_{n'm'}(p)=T_{nm}(p), \tag{17}$$

where

$$T_{nm}(p)=\frac{1}{2\pi\mu a}\int_0^a dz\int_{-\infty}^{\infty}dy\;t_n^i(z,y)\phi_m(z)e^{ipy}, \tag{18}$$

and

$$Q_{nmn'm'}(p)=\frac{\alpha_{nn'}}{\mu k_s a}\int_0^a\phi_m(z)\phi_{m'}(z)dz+\frac{4\pi i}{a}(-1)^{m'}\int_{-\infty}^{\infty}\frac{dq}{q^2}\sum_j \frac{1}{k_j h_j}F_{jn'}F_{jn}^*J_m(qa/2)J_{m'}(qa/2) - \frac{4\pi i}{a}\int_{-\infty}^{\infty}dq\sum_{jj'}\frac{1}{k_j h_j}G_{jn'}R_{jj'}G_{j'n}^+\exp[i(h_j+h_{j'})(b+\cos\psi\;a/2)] \times\frac{J_m((h_j\cos\psi+q\sin\psi)a/2)}{h_j\cos\psi+q\sin\psi}\,\frac{J_{m'}((h_{j'}\cos\psi-q\sin\psi)a/2)}{h_{j'}\cos\psi-q\sin\psi}. \tag{19}$$

The limit $x' \to 0$ in front of the integral in Eq. (3) has been moved inside the integrals because the integrals in Eq. (19) are convergent due to the asymptotic behavior in Eq. (13) and the asymptotic behavior of the Bessel functions. The first integrand goes like q^{-1} times the two Bessel functions and this slow convergence is not suitable for numerical integration. The leading order term can be integrated analytically, however, and for the rest the convergence is quick enough for a direct numerical integration. The second integral in Eq. (19) decays exponentially for $b \neq 0$, but if $b = 0$ the integrand goes like q^{-2}. This leading part is also integrated analytically. To avoid the cuts and the Rayleigh pole along the real q axis, the integration contour is deformed a little into the second and fourth quadrants.

Once the COD expansion coefficients $\beta_{nm}(p)$ are determined from Eq. (17), it is in principle straightforward to obtain the scattered field from the integral representation Eq. (2). However, a problem is that the Fourier transform variable p in Eq. (17) is a parameter in the equation. If the situation is such that a continuous set of p values is needed, then the numerical procedure will become very costly. This will be avoided here by assuming that the crack is in the far field from an ultrasonic probe, whereby a stationary-phase analysis can be performed so that only the value $p = 0$ is needed.

II. THE INCOMING PROBE FIELD

In the previous section the incoming field was left arbitrary, but in this section it is specified as the field from an ultrasonic contact probe. In the next section the modelling of the field received by another (possibly the same) probe is considered. The total geometry with transmitting and receiving probes, the striplike crack and the free back surface is depicted in Fig. 2. The two probes are located on the same planar surface. At each probe a local coordinate system $x_t y_t z_t$ and $x_r y_r z_r$ (the suffices t and r are attached to quantities referring to the transmitting and receiving probes, respectively) is fixed with the origin at the index point of the probe, the z_t and z_r axes normal to the scanning surface, and the y_t and y_r axes parallel to the y and y_0 axes, i.e., parallel to the length of the crack. The x_t and x_r axes are assumed to coincide, i.e., the line between the two probes is perpendicular to the length of the crack. The free surface close to the crack can be tilted around the y axis with the angle η, which means that the crack is tilted $\eta + \psi$. If $\eta \neq 0$ the two planar surfaces will of course intersect, but this intersection is assumed to lie far away from the probes and the crack, so it is of no importance. All multiple scattering between the scanning surface and the crack or back surface is neglected. This is a good approximation as long as the crack is some distance away from the scanning surface. The vectors from the probes to the lower crack edge are denoted $\mathbf{d}_t$ and $\mathbf{d}_r$ and are located in the xz plane.

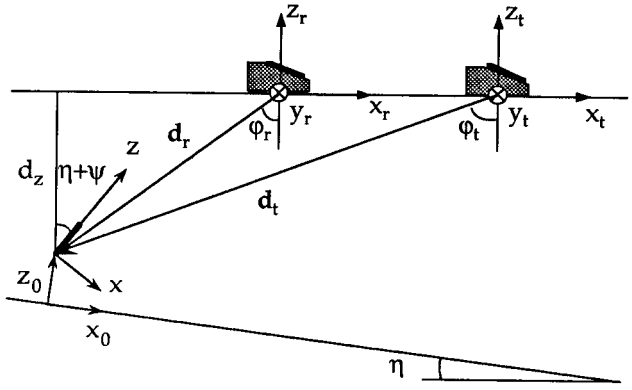


FIG. 2. The geometry of the transmitting and receiving probes, the crack, and the free back surface.

The modeling of the transmitting probe is described by Boström and Wirdelius.[10] The action of the probe is modelled by the traction beneath the probe. Both the normal and tangential traction components can be included. The effective contact area of the probe can be elliptic or rectangular and the probe can be of any type (SH, SV, or P) and arbitrarily angled. To solve the boundary-value problem with a given traction vector on the surface of an elastic half-space a double Fourier transform in the coordinates parallel to the surface is used. The result is that the radiated field from the probe is

$$u_n(\mathbf{r}_t) = \sum_j \int\int \frac{dq_t\, dp_t}{k_j h_{tj}}\, \xi_j^t(q_t, p_t) g_{jn}^*(q_t, p_t) \times e^{i(q_t x_t + p_t y_t - h_{tj} z_t)}. \tag{20}$$

Here $h_{tj} = \sqrt{k_j^2 - q_t^2 - p_t^2}$ with $\mathrm{Im}\, h_{tj} \geqslant 0$ and g_{jn}^* is given by Eq. (6) with q, p, h_j exchanged with q_t, p_t, $-h_{tj}$. The coefficients ξ_j^t are given by Boström and Wirdelius[10] and it is noted they can be given explicitly in terms of elementary or Bessel functions. The q_t and p_t integration intervals are the whole real axis. To avoid problems below, only those parts of the integration intervals are kept that belong to propagating or down-going waves in both the $x_t y_t z_t$ and $x_0 y_0 z_0$ systems. In the next section a stationary-phase analysis will anyway pick up only a single value of q_t and p_t.

The direct probe field in Eq. (20) does not satisfy the boundary condition on the free back surface as required by the incoming field from the previous section. To add the field reflected from the back surface it is convenient to first transform the expansion in Eq. (20) which is expressed in the $x_t y_t z_t$ system to the $x_0 y_0 z_0$ system. This can be done very systematically with the formalism developed by Boström *et al.*[24] In principle this formalism is employed here, but the formulas are written a little more explicitly.

The translation from the $x_t y_t z_t$ system to the $x_0 y_0 z_0$ system is done with the vector $\mathbf{d}_t - b\hat{z}_0$ (where $\hat{z}_0$ is the unit vector in the z_0 direction). A rotation with the angle η then transforms the plane waves to the $x_0 y_0 z_0$ system:

$$u_n(\mathbf{r}_0) = \sum_{jj'} \int\int \frac{dq_t\, dp_t}{k_j h_{tj}}\, \xi_j^t(q_t, p_t) \times e^{id_t(-q_t \sin\varphi_t + h_{tj}\cos\varphi_t) + ibh_{0j'}} \times \Re_{jj'}(-\eta) g_{j'n}^*(q_{0j}, p_0) e^{i(q_{0j'}x_0 + p_0 y_0 - h_{0jj'}z_0)}. \tag{21}$$

Here $\mathfrak{R}_{jj'}$ are (the elements of) the rotation matrix

$$\mathfrak{R}_{jj'}=\begin{pmatrix} \cos\Omega & -i\sin\Omega & 0 \\ -i\sin\Omega & \cos\Omega & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{22}$$

where Ω depends on q_t, p_t, and η in a rather complicated manner[24] (but only $\Omega=0$ or π will enter later in the calculations). The wave numbers q_{0j}, p_0, $h_{0jj'}$ depend on the integration variables and the rotation η so that

$$q_{0j}=q_t\cos\eta+h_{tj}\sin\eta, \tag{23a}$$

$$p_0=p_t, \tag{23b}$$

$$h_{0jj'}=\sqrt{k_{j'}^2-q_{0j}^2-p_0^2}, \quad \mathrm{Im}\, h_{0jj'}\geqslant 0. \tag{23c}$$

As seen by the notation q_{0j} and $h_{0jj'}$ depend on j (and j' for $h_{0jj'}$; this is one reason why it is in some respects more natural to use the spherical angles of propagation of plane waves rather than rectangular components[24]).

When the probe field is written as in Eq. (21) with the plane waves in the $x_0y_0z_0$ system it is of course clear how to add a reflected wave so as to get a total incoming field that satisfies the boundary condition on the free surface. The total incoming traction field on the crack ($x=0$) and transformed to the crack system is

$$\begin{aligned} t_n^i(z,y)=\sum_{jj'}\int\int \frac{dq_t\,dp_t}{k_jh_{tj}}\,\xi_j^t(q_t,p_t)e^{id_t(-q_t\sin\varphi_t+h_{tj}\cos\varphi_t)} \\ \times\Bigg[\mathfrak{R}_{jj'}(-\eta-\psi)G_{j'n}^*(q_j,p;\psi=0)e^{i(py-h_{jj'}z)} \\ +\sum_{j''}\mathfrak{R}_{jj'}(-\eta)R_{j''j'}(q_{0j},p_0)G_{j''n}(q_{0j},p_0) \\ \times e^{i(p_0y+z(q_{0j}\sin\psi+h_{0jj''}\cos\psi)+b(h_{0jj'}+h_{0jj''}))}\Bigg]. \end{aligned} \tag{24}$$

The first term is the direct field from Eq. (2) but transformed to the xyz system and the second term is the field reflected from the free back surface. The quantities $G_{j''n}(q_{0j},p_0)$ are given in Eq. (12) with q_{0j} and p_0 substituted for q and p and $G_{j'n}^*(q_j,p;\psi=0)$ are likewise given by Eq. (12) with q_j substituted for q and $\psi=0$. The free-surface reflection coefficients $R_{j''j'}(q_{0j},p_0)$ are given in Eq. (8) with q_{0j} and p_0 substituted for q and p. The wave numbers q_j, p and $h_{jj'}$ in the crack system depend on the integration variables and the rotation $\eta+\psi$ so that

$$q_j=q_t\cos(\eta+\psi)+h_{tj}\sin(\eta+\psi), \tag{25a}$$

$$p=p_t, \tag{25b}$$

$$h_{jj'}=\sqrt{k_{j'}^2-q_j^2-p^2}, \quad \mathrm{Im}\, h_{jj'}\geqslant 0. \tag{25c}$$

The total incoming traction field given in Eq. (24) is thus the field to be inserted into the source term defined in Eq. (18) of the integral equation.

III. THE SIGNAL RESPONSE OF THE RECEIVING PROBE

Consider now the geometry of Fig. 2 with both a transmitting and a receiving probe (as a special case they may of course coincide). To model the electrical signal response in the receiving probe, the reciprocity argument of Auld[25] is ideal to use as it only presupposes a knowledge of the probe acting as transmitter. According to Auld the signal response is obtained as an integral over the crack

$$\delta\Gamma(\omega)=-\frac{i\omega}{4P}\int_{-\infty}^{\infty}dy\int_0^a dz\;\Delta\mathbf{u}\cdot\mathbf{t}_1, \tag{26}$$

where $\delta\Gamma$ is the change in the electrical reflection coefficient of the receiving probe, i.e., more or less the quantity measured in practice. The COD $\Delta\mathbf{u}$ is the one calculated in previous sections and $\mathbf{t}_1$ is the traction on the crack which would exist if the receiver acts as a transmitter in the absence of the crack. The same electric effect P is incident on both probes when acting as transmitters. Note that the result is valid at a fixed frequency and that all materials are assumed lossless. Also note that the result is convenient to apply as only the COD and not the scattered field is required.

When the expansion of the COD from Eq. (14) and the traction from the receiver from Eq. (24) (with the receiver characteristics inserted instead of the transmitter ones) the y and z integrals in Eq. (26) and subsequently the p' integral in Eq. (14) can be performed and the result can be written as

$$\begin{aligned} \delta\Gamma=-\frac{i\omega\mu}{4Pk_s}2\pi\sum_{nm}\sum_{jj'}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{dq_r\,dp_r}{k_jh_{rj}}\,\xi_j^r(q_r,p_r) \\ \times e^{id_r(-q_r\sin\varphi_r+h_{rj}\cos\varphi_r)}\beta_{nm}(p_r) \\ \times[\mathfrak{R}_{jj'}(-\eta-\psi)M_{jj'nm}(q_j) \\ +\mathfrak{R}_{jj'}(-\eta)N_{jj'nm}(q_{0j})], \end{aligned} \tag{27}$$

where

$$\begin{aligned} M_{jj'nm}(q_j)=(-1)^{m+1}G_{j'n}^*(q_j,p,\psi=0) \\ \times\frac{1}{h_{jj'}}\,e^{-ih_{jj'}a/2}J_m(h_{jj'}a/2), \end{aligned} \tag{28}$$

$$\begin{aligned} N_{jj'nm}(q_{0j})=\sum_{j''}e^{ib(h_{0jj'}+h_{0jj''})}R_{j''j'}(q_{0j},p_0)G_{j''n}(q_{0j},p_0) \\ \times\frac{1}{O_{jj''}}\,e^{iO_{jj''}a/2}J_m(O_{jj''}a/2), \end{aligned} \tag{29}$$

$$O_{jj''}=q_{0j}\sin\psi+h_{0jj''}\cos\psi. \tag{30}$$

The p_r dependence is suppressed in $N_{jj'nm}$ and $M_{jj'nm}$. The various q and p variables are connected as described in the previous section with appropriate changes of index t to index r.

As it stands Eq. (27) is inconvenient to use because it contains a double integral over infinite ranges. Furthermore, the expansion coefficients $\beta_{nm}(p_r)$ that is the solution of Eq. (17) have to be solved for all p_r values that are needed in the integral. Here an approximate evaluation of the integrals by a repeated application of the stationary-phase approximation is

employed to circumvent the difficulties with the double integral. A necessary condition for the application of this approximation is that the distance between the probes and the crack are many wavelengths. As the multiple scattering between the scanning surface and the crack has already been neglected this condition has already been implicitly assumed. Further conditions on the stationary phase approximation are discussed below.

The large parameter in the p_r integral in Eq. (27) is the crack depth $d_z = d_r \cos\varphi_r = d_t \cos\varphi_t$ (normalized with k_s or k_p). The phase function is

$$\psi = h_{rj} + h_{tj'} = \sqrt{k_j^2 - q_r^2 - p_r^2} + \sqrt{k_{j'}^2 - q_t^2 - p_r^2}. \tag{31}$$

The first term comes from the exponential in Eq. (27) and the second comes via $\beta_{nm}(p_r)$ and $T_{nm}(p_r)$; see Eqs. (17) and (18) [with Eq. (24) inserted, Eq. (18) is similar to Eq. (27), the changes being that the p_r integral and the factor $\beta_{nm}(p_r)$ are missing and that ''r'' is exchanged to ''t'']. The stationary point of ψ is immediately seen to be at $p_r = 0$. It is then easy to calculate ψ and ψ'' at $p_r = 0$ and hereby obtain the stationary-phase approximation of the p_r integral in Eq. (27).

The q_r integral in Eq. (27) is also of stationary-phase type with large parameter d_r and phase function

$$\psi = -q_r \sin\varphi_r + \sqrt{k_j^2 - q_r^2}. \tag{32}$$

The stationary point is then

$$q_r = -k_j \sin\varphi_r \tag{33}$$

and at this point it is easy to calculate

$$\psi = -k_j, \tag{34}$$

$$\psi'' = -\frac{1}{k_j \cos^2\varphi_r}. \tag{35}$$

The q_t integral remaining in Eq. (18) is of exactly the same type and thus has stationary point at $q_t = -k_j \sin\varphi_t$ and $\psi = -k_j$ and $\psi'' = -1/k_j \cos^2\varphi_t$ at this point.

With $p_r = 0$ the rotation matrix $\Re_{jj''}$ of Eq. (22) becomes diagonal with $\cos\Omega = \pm 1$. Collecting everything and simplifying finally gives the signal response as

$$\begin{aligned}\delta\Gamma = &-\frac{\mu c_s}{4P} e^{-i\pi/4} \frac{(2\pi)^{5/2}}{a\sqrt{d_z d_r d_t}} \sum_{jj'} \frac{e^{i(k_j d_r + k_{j'} d_t)}}{k_j k_{j'}} \sqrt{\frac{\cos\varphi_r \cos\varphi_t}{k_j \cos\varphi_r + k_{j'} \cos\varphi_t}}\, \xi_j^r(-k_j \sin\varphi_r, 0)\, \xi_{j'}^t(-k_{j'} \sin\varphi_t, 0) \\ &\times \sum_{nm} \sum_{n'm'} (Q^{-1})_{nmn'm'} [\Omega_j(\varphi_r, \varphi_r - \eta - \psi) M_{jjnm}(-k_j \sin(\varphi_r - \eta - \psi), 0) \\ &+ \Omega_j(\varphi_r, \varphi_r - \eta) N_{jjnm}(-k_j \sin(\varphi_r - \eta), 0)][\Omega_{j'}(\varphi_t, \varphi_t - \eta - \psi) M_{j'j'n'm'}(-k_j \sin(\varphi_t - \eta - \psi), 0) \\ &+ \Omega_{j'}(\varphi_t, \varphi_t - \eta) N_{j'j'n'm'}(-k_{j'} \sin(\varphi_t - \eta), 0)],\end{aligned} \tag{36}$$

where

$$\Omega_j(\varphi, \theta) = \begin{cases} \text{sign}(\sin\varphi \sin\theta), & j = 1,2, \\ 1, & j = 3. \end{cases} \tag{37}$$

This is a reasonably compact expression which is straightforward to compute. By far the most complicated and time-consuming part is the matrix $Q_{nmn'm'}$ given in Eq. (19).

When performing the stationary-phase approximation it is assumed that $k_j d_z$ is large, i.e., that the crack depth is large. For the validity of the approximation all other factors in the integrands must be slowly varying in the vicinity of the stationary point. The probe functions ξ^r and ξ^t contain factors of the type $\sin qc$ with c the radius of the probe. For this to be slowly varying c must be small compared to the crack depth d_z. Furthermore, the probe near field length c^2/λ, where λ is the (P or S, depending on probe type) wavelength, must be smaller than the depth d_z and this condition is the one most often violated in practical applications. The factors in the integrand originating from the crack contain similar oscillating factors and thus the crack width must be much smaller than the depth d_z and $a^2/4\lambda$ should be smaller than d_z. All in all there are several conditions that must be met for the validity of the stationary phase approximations. In most practical applications these conditions are more or less satisfied.

How accurate the stationary-phase approximation is can be investigated by changing the large parameter in the calculation. As seen from the calculations it is the vectors $\mathbf{d}_t$ and $\mathbf{d}_r$ that act as large parameters and these are somewhat arbitrarily drawn from the center of the probe to the lower crack edge. Alternatives would be to draw them to the upper crack edge or the centre of the crack. The only changes in the stationary-phase calculations is to change $\mathbf{d}_r$ and $\mathbf{d}_t$ and compensate with a phase factor. The use of the alternatives is further commented in the next section.

IV. NUMERICAL EXAMPLES

In this section a few numerical examples comparing the results from the present theory with other calculations and experiments are given.

The numerical calculations are fairly straightforward from the formulas given in the preceding sections. The only tricky point is the computations of the infinite integrals in Eq. (19) which were commented on in Sec. I. The m and n indices in Eq. (36) run from 1 to ∞. It is found that a suitable truncation is to take the largest m value as $k_s a/2 + 10$. As experiments are performed in the time domain all the results

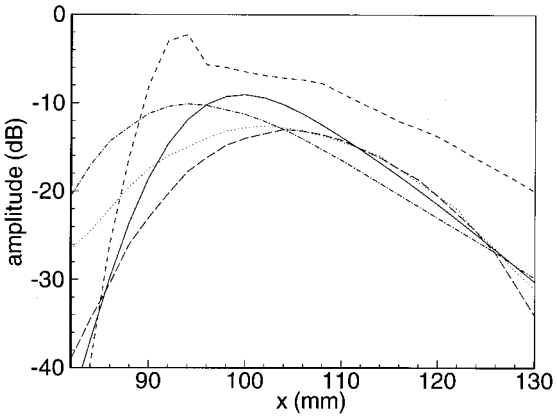


FIG. 3. The signal response from the lower crack edge as a function of probe position for a vertical crack of width 10 mm at center depth 55 mm. Comparison between experiment and four theories; ——— UTDefect, ... CISE experiment, – – – – – – NDTAC, - - - - - - - - - TOFD, and –·–·–·– Kassel.

in the following are also computed in the time domain. This is accomplished by taking an inverse Fourier transform numerically using the frequency spectrum.

$$S(f)=\begin{cases}\dfrac{1}{\Delta f}\cos^2\dfrac{\pi(f-f_c)}{2\Delta f}, & f_c-\Delta f<f<f_c+\Delta f,\\ 0, \quad \text{otherwise},\end{cases} \tag{38}$$

where f_c is the center frequency and Δf the 6-dB bandwidth. Usually 100 frequencies equally spaced are enough to obtain accurate results.

The first few results are for an internal striplike open crack, where the results can be compared with data from the PISC project, cf. Lakestani.[26] An internal crack without any back surface can of course be obtained as a special case simply by omitting the last term in Eq. (19) and the two N terms in Eq. (36). Figures 3 and 4 show the maximum amplitude as a function of probe position for a 10-mm vertical crack at center depth 55 mm. The same probe is used both as transmitter and receiver. It is a 60° SV probe of rectangular crystal size 20 mm by 22 mm which has a center frequency

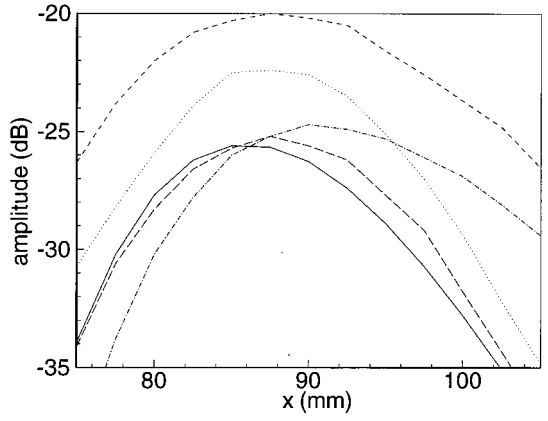


FIG. 4. Same as Fig. 3 but for upper crack edge.

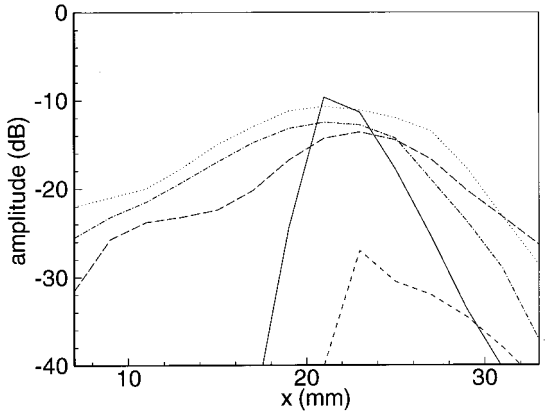


FIG. 5. Same as Fig. 3 but for crack at center depth 8 mm.

of 2.04 MHz and a bandwidth of 2.04 MHz. In this case the dominating scattering mechanism is edge diffraction and Fig. 3 therefore shows the diffraction from the lower crack edge and Fig. 4 that from the upper crack edge. The two edge diffractions are of course identified by their different arrival times (for this reason it is essential to perform the computations in the time domain). The figures are individually normalized by a 9.5-mm-diameter side-drilled hole with its center at the same depth as the respective crack edge (the scattering of a probe field by a cylindrical cavity is performed in a similar way as in the present paper[27]). Five curves are shown in the figures: the present calculations (marked UTDefect), an experimental curve (marked CISE), a refined GTD curve (marked NDTAC), a regular GTD (marked TOFD), and a two-dimensional finite integration computation (marked Kassel). Lakestani[26] describes the methods and gives some references. Comparing the different curves it is seen that they all are similar giving peak values and -6 dB widths within a few dB and peak position within 10 mm. For these cases the TOFD curves are the ones that are most dissimilar.

There may be many sources for the differences between the curves: experimental errors, different uncontrolled probe parameters, such as frequency spectra, angle, size, and coupling conditions, and mathematical approximations. The models also neglect damping and assume that the crack is thin, open, without roughness and with absolutely sharp edges. For the present computations the most serious limitation is the use of the stationary phase method. As discussed in the previous section this in particular requires that the distance between the probe and the crack is greater than the probe's near field length. The near field length of the probe used in Figs. 3 and 4 is 76 mm so this condition is not quite satisfied. As mentioned in the previous section, there is a possibility to check the stationary-phase approximation by changing the large parameter. In Fig. 3 the large parameter is to the lower crack edge as in the equations of the previous section. This gives the peak value -9.1 dB at $x=100$ mm. Changing the large parameter to the upper crack edge instead gives -6.7 dB at $x=83$ mm. Particularly the location of the peak changes significantly. Figure 4 is instead drawn with the large parameter to the upper crack edge giving a peak

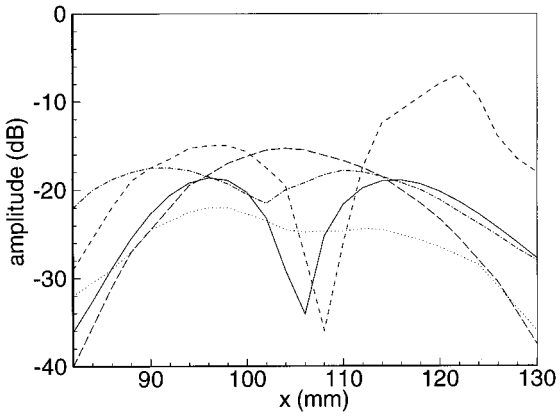


FIG. 6. Same as Fig. 3 but for crack tilted 7° toward the probe.

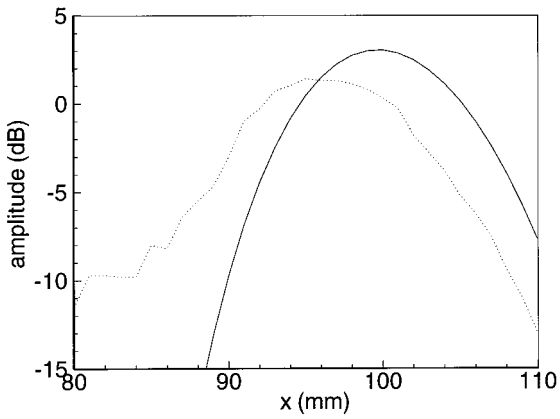


FIG. 8. Same as Fig. 7 but crack tilted 10° away from the probe.

value -25 dB at $x=85$ mm. Changing the large parameter to the lower edge instead gives -27.9 at $x=103$ mm, again changing the location more than the peak value. The conclusion is that the stationary-phase approximation performs rather poorly for the conditions of Figs. 3 and 4, but that acceptable values can be obtained by choosing the large parameter so that it goes to the diffraction point of interest.

Another comparison with the PISC project is shown in Fig. 5 which shows the same type of curves for a 10-mm vertical crack with center depth 8 mm. It is only the diffraction echo from the lower crack edge that is shown and the calibration is again performed with a side-drilled hole at the same center depth as the lower crack edge. The same probe is used as in Figs. 3 and 4. The height and the location of the peak is well predicted by UTDefect, but the width is very much smaller than for the experiment. The reason for this is that the crack is very close to the scanning surface within the probe's near field. The stationary-phase approximation is not valid in this case. The stationary-phase approximation gives the probe field as if it originated from one point (a spherical wave) and in the near field this greatly underestimates the width of the probe field thus leading to the too small width in Fig. 5. The strength of the probe field on the beam axis seems to be well predicted even in the near field, thus giving a reasonable height of the UTDefect curve in Fig. 5.

A final comparison with the PISC project is given in Fig. 6 which is identical with Fig. 3 except that the crack is tilted 7° toward the probe. This gives an incident field on the crack close to the critical angle and therefore the differences with Fig. 3 are significant. Also in this case UTDefect gives results that are reasonably close to experiments and the other methods.

From a practical point of view the most interesting crack is the surface-breaking one. Strictly speaking the present formalism is not valid in this case as the correct edge condition is not satisfied. The Chebyshew functions in Eq. (15) will in fact force the crack mouth to be closed whereas an opening is of course to be expected in reality. However, for a corner echo from a surface-breaking crack which is not too small (one or two wavelengths at least) the exact crack edge behavior should be of minor importance as the specular reflections dominate. Thus reasonable results could be expected also for surface-breaking cracks and the following comparisons with experiments seem to confirm this.

Figures 7 and 8 show the maximum amplitude as a function of probe position for an 8-mm surface-breaking crack in

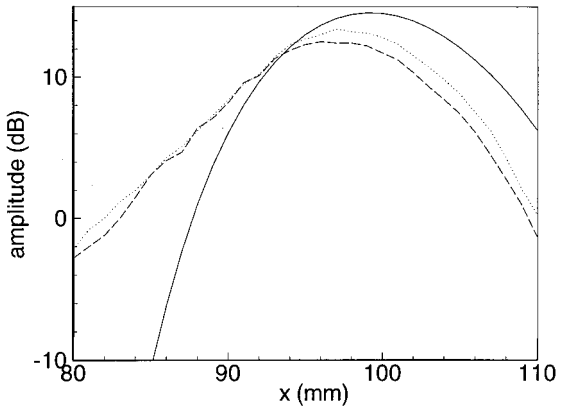


FIG. 7. The signal response as a function of probe position for a vertical surface-breaking crack of width 8 mm in a 100-mm-thick steel plate. Comparison with experiments; ——— UTDefect, ... and – – – – – – experiments from two directions.

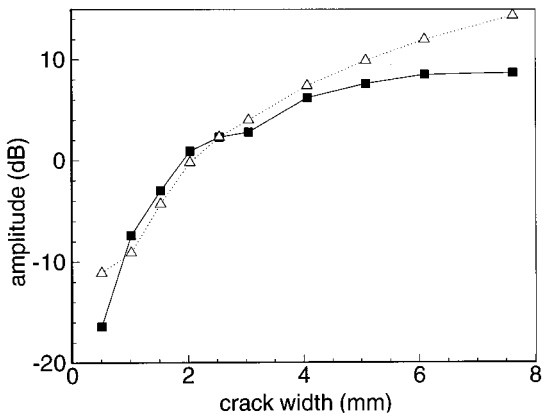


FIG. 9. Maximum signal response as a function of crack width for a surface-breaking crack in a 15.24-mm-thick steel plate; ——— experiments, ... UTDefect.

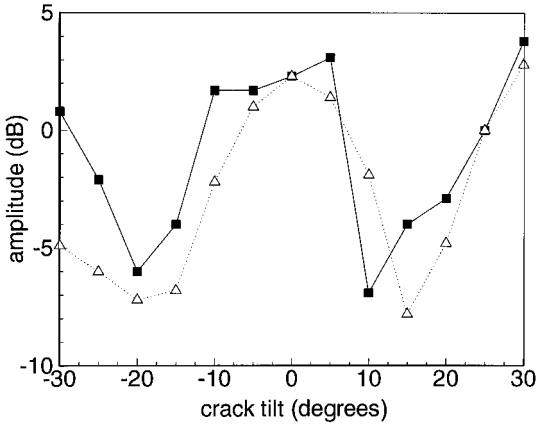


FIG. 10. Maximum signal response as a function of crack tilt for a surface-breaking crack of width 2.54 mm in a 15.24-mm-thick steel plate; ——— experiments, ... UTDefect.

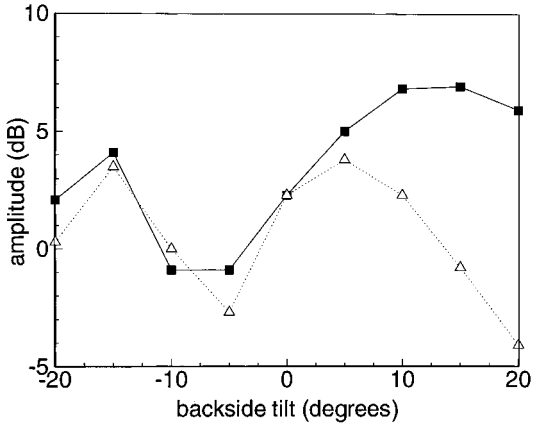


FIG. 11. Maximum signal response as a function of backside tilt for a surface-breaking crack of width 2.54 mm in a 15.24-mm-thick steel plate; ——— experiments, ... UTDefect.

100-mm-thick steel plate. Figure 7 is for a vertical crack whereas Fig. 8 is for a crack tilted $-10°$, i.e., the corner seen from the probe is 100°. A single probe is used as both transmitter and receiver. It is a 45° SV probe of rectangular crystal size 20 mm by 22 mm which has a center frequency of 2.1 MHz and a bandwidth of 2.1 MHz. The calibration is performed by a side-drilled hole of diameter 6.5 mm at the center depth 80 mm. The experiments have been performed at AEA in England (under contract to Nuclear Electric, Chapman[28]). In Fig. 7 there are two experimental curves, one taken from each side of the crack, whereas Fig. 8 only contains one experimental curve. The comparison between UTDefect and experiment is seen to be good. The peak values are differing 1–2 dB and the peak position is about 2–3 mm off. Outside the peaks the experimental and UTDefect curves fall off a little differently. As opposed to peak value and position, this behaviour is sensitive to the probe size, frequency, and bandwidth. In this case the bandwidth and frequency spectrum is not very well known for the experiments and the choice of the effective size of the probe is somewhat arbitrary, so it is not surprising that discrepancies show up. Experimental results are also available for crack tilt 2°, 4°, 6°, and 8° and comparing these with UTDefect show a similar agreement with peak values at most 2 dB off.

Becker *et al.*[29] perform experiments with surface-breaking cracks, tilting the crack or the backside, and using four different probe sizes. Only comparisons with the smallest probe is performed, this being a 2.25-MHz circular SV probe of diameter 6.35 mm. As the steel plate containing the cracks is 15.24 mm thick, this probe operates outside its near field length of 7 mm. The other probes used by Becker *et al.*[29] are at least twice as big and thus operates within their near field. As discussed previously this leads to problems with the validity of the stationary-phase approximation.

Comparisons are only performed for a 60° SV probe. The bandwidth is unknown but is taken as 2 MHz. Figure 9 shows the maximum response as a function of crack width for widths between 0.508 mm and 7.602 mm (Becker *et al.*[29] are working with inches). The UTDefect results are uncalibrated; instead the curve is adjusted so that the result for the 2.54-mm crack coincides with the experimental value. Except for the values for the largest and smallest widths the results are seen to agree very well with discrepancies within about 1 dB. For the largest cracks the stationary-phase approximation may become uncertain, but the reason for the difference between the experimental and UTDefect results for the smallest width is harder to understand.

Figure 10 shows a comparison between UTDefect and Becker *et al.*[29] for a crack of width 2.54 mm but with varying tilt between $\pm 30°$. The UTDefect results are again normalized so that the 0° crack result agrees with the experiment. The trends of the two curves are similar. For eight of the points the difference between the curves are less than 2 dB. For four of the other points the differences are about 4 dB and for one point $(-30°)$ it is about 6 dB.

A final comparison with Becker *et al.*[29] is given in Fig. 11, where the crack is vertical of width 2.54 mm but the tilt of the back side is varied. Again the UTDefect results are normalized so that the untilted crack result agrees with the corresponding experimental value. For angles between $-20°$ and $+5°$ the agreement is very good, but for increasing positive tilts the differences become progressively larger with almost 10 dB at $+20°$. The reason for these discrepancies are unknown but from an intuitive point of view the UTDefect results with a decreasing response with increasing tilt seem more reasonable.

ACKNOWLEDGMENT

The present work is sponsored by the Swedish Nuclear Power Inspectorate (SK) and this is gratefully acknowledged.

[1] R. K. Chapman, ''A system model for the ultrasonic inspection of smooth planar cracks,'' J. Nondestruct. Eval. **9**, 197–211 (1990).
[2] J. M. Coffey and R. K. Chapman, ''Application of elastic scattering theory for smooth flat cracks to the quantitative prediction of ultrasonic defect detection and sizing,'' Nucl. Energy **22**, 319–333 (1983).
[3] P. Fellinger, R. Marklein, K. J. Langenberg, and S. Klaholz, ''Numerical modeling of elastic wave propagation and scattering with EFIT-elastodynamic finite integration technique,'' Wave Motion **21**, 47–66 (1995).
[4] W. Lord, R. Ludwig, and Z. You, ''Developments in ultrasonic modeling with finite element analysis,'' J. Nondestruct. Eval. **9**, 129–143 (1990).

[5] K. Harumi and M. Uchida, ''Computer simulation of ultrasonics and its applications,'' J. Nondestruct. Eval. **9**, 81–99 (1990).
[6] B. Nouailhas, G. Van Chi Nguyen, F. Pons, and S. Vermersch, ''Ultrasonic modeling and experiments: an industrial case: bimetallic weld in nuclear power plant,'' J. Nondestruct. Eval. **9**, 145–153 (1990).
[7] A. Minachi, J. Mould, and R. B. Thompson, ''Ultrasonic beam propagation through a bimetallic weld—a comparison of predictions of the Gauss–Hermite beam model and finite element model,'' J. Nondestruct. Eval. **12**, 151–158 (1993).
[8] T. Lawn, M. Veda, and M. Tabei, ''Ultrasonic scattering from a simulated cavity in steel,'' J. Acoust. Soc. Am. **98**, 2809–2818 (1995).
[9] A. Lhémery, ''Impulse-response method to predict echo responses from defects in solids. Part 1. Theory,'' J. Acoust. Soc. Am. **98**, 2197–2208 (1995).
[10] A. Boström and H. Wirdelius, ''Ultrasonic probe modeling and nondestructive crack detection,'' J. Acoust. Soc. Am. **97**, 2836–2848 (1995).
[11] K. J. Langenberg, P. Fellinger, R. Marklein, P. Zanger, K. Meyer, and T. Kreutler, ''Inverse methods and imaging,'' in *Evaluation of Materials and Structures by Quantitative Ultrasonics*, edited by J. D. Achenbach (Springer-Verlag, Vienna, 1993).
[12] R. J. Brind and J. D. Achenbach, ''Scattering of longitudinal and transverse waves by a sub-surface crack,'' J. Sound Vib. **78**, 555–563 (1981).
[13] Y. C. Angel and J. D. Achenbach, ''Reflection and transmission of obliquely incident Rayleigh waves by a surface-breaking crack,'' J. Acoust. Soc. Am. **75**, 313–319 (1984).
[14] V. K. Kinra and B. Q. Vu, ''Diffraction of Rayleigh waves in a half-space. II. Inclined edge crack,'' J. Acoust. Soc. Am. **79**, 1688–1692 (1986).
[15] N. Saffari and L. J. Bond, ''Body to Rayleigh wave mode-conversion at steps and slots,'' J. Nondestruct. Eval. **6**, 1–22 (1987).
[16] C. L. Scandrett and J. D. Achenbach, ''Time-domain finite difference calculations for interaction of an ultrasonic wave with a surfacebreaking crack,'' Wave Motion **9**, 171–190 (1987).
[17] Ch. Zhang and J. D. Achenbach, ''Scattering of body waves by an inclined surface-breaking crack,'' Ultrasonics **26**, 132–138 (1988).
[18] W. Lin and L. M. Kerr, ''Scattering by horizontal subsurface pennyshaped crack,'' Proc. R. Soc. London, Ser. A **408**, 227–294 (1986).
[19] Ch. Zhang and J. D. Achenbach, ''Dispersion and attenuation of surface waves due to distributed surface-breaking cracks,'' J. Acoust. Soc. Am. **88**, 1986–1992 (1990).
[20] L. Guan and A. Norris, ''Elastic wave scattering by rectangular cracks,'' Int. J. Solids Struct. **29**, 1549–1565 (1992).
[21] S. Krenk and H. Schmidt, ''Elastic wave scattering by a circular crack,'' Philos. Trans. R. Soc. London, Ser. A **308**, 167–198 (1982).
[22] A. Boström, ''Acoustic scattering by a sound-hard rectangle,'' J. Acoust. Soc. Am. **90**, 3344–3347 (1991).
[23] P. A. Martin and F. J. Rizzo, ''On boundary integral equations for crack problems,'' Proc. R. Soc. London, Ser. A **421**, 341–355 (1989).
[24] A. Boström, G. Kristensson, and S. Ström, ''Transformation properties of plane, spherical and cylindrical scalar and vector wave functions,'' in *Field Representations and Introduction to Scattering*, edited by V. V. Varadan, A. Lakhtakia, and V. K. Varadan (Elsevier, Amsterdam, 1991).
[25] B. A. Auld, ''General electromechanical reciprocity relations applied to the calculation of elastic wave scattering coefficients,'' Wave Motion **1**, 3–10 (1979).
[26] F. Lakestani, ''Validation of mathematical models of the ultrasonic inspection of steel components,'' PISC III rep. 16, JRC, Inst. Adv. Mat. Petten, The Netherlands (1992).
[27] A. Boström and P. Bövik, report to be published.
[28] R. K. Chapman (private communication).
[29] F. L. Becker, S. R. Doctor, P. G. Heasler, C. J. Morris, S. G. Pitman, G. P. Selby, and F. A. Simonen, ''Integration of NDE reliability and fracture mechanics,'' NURE/CR 1696, PNL-3469, VR. 1, U.S. Nuclear Regulatory Commission (1981).

Acoustic field prediction for a single planar continuous-wave source using an equivalent phased array method

Xiaobing Fan, Eduardo G. Moros, and William L. Straube
Radiation Oncology Center, Washington University School of Medicine, St. Louis, Missouri 63108

(Received 19 February 1997; accepted for publication 25 July 1997)

Phased array theory is combined with the Rayleigh–Sommerfeld diffraction integral to predict measured acoustic fields generated by a single-source ultrasonic transducer. The idea is to treat a single-source as a ''phased array,'' which is composed of many small elements. The goal is to find the excitation source for the phased array, that is, the amplitude and phase for each array element, which produces an acoustic field similar to the experimentally measured field generated by the single-source transducer. A pressure field measured at a given plane parallel and close to the face of the transducer in degased water was used to calculate the excitation source of the equivalent phased array using an inverse technique. The excitation source of the equivalent phased array was then used to calculate the acoustic field from this measurement plane to the far field. It was demonstrated that this phased array approach accurately predicted the location of major grating lobes and the general distributions of the near and far pressure fields for four different transducers. This equivalent phased array method (EPAM) used to model a single-source transducer should be useful in both diagnostic and therapeutic ultrasound applications. © *1997 Acoustical Society of America.* [S0001-4966(97)03111-1]

PACS numbers: 43.38.Ar, 43.35.Yb [SLE]

INTRODUCTION

Theoretical studies of acoustic fields generated by a circular or rectangular piston transducer vibrating within an infinite plane rigid baffle have been investigated and well-characterized for both continuous and transient excitations (Freedman 1960; Zemanek, 1971; Lockwood and Willette, 1973; Marini and Rivenez, 1974; Weyns, 1980; Hutchins *et al.*, 1986, 1987; Ocheltree and Frizzell, 1989). The Rayleigh–Sommerfeld diffraction integral is used to calculate acoustic fields for a plane piston. Several calculation methods have been developed to evaluate this integral by dividing the radiating surface into elements (Stepanishen, 1971; Zemanek, 1971; Ocheltere and Frizzell, 1989), or using the Fourier transform approach (Lockwood and Willette, 1973; Williams and Maynard, 1982). Most theoretical calculations have been performed under the assumption that particle velocities on the surface of a transducer have uniform amplitude and phase distributions. However, a theoretical model that calculates the acoustic field for a planar transducer using a uniform particle velocity distribution does not produce good agreement with experimental results. For example, Hutchins *et al.* (1986) found that lead zirconate-titanate (PZT) transducers had complicated variations in amplitude across their faces. This could be one reason that theoretical and experimental results have not been in good agreement. Clearly, the variation of particle velocity distributions on the surface of a transducer has an impact on the acoustic field distributions that it produces.

The particle velocity on the surface of an ultrasound transducer is difficult to determine. Greenspan (1979) examined the acoustic field generated from a transducer with a Gaussian particle velocity distribution across its face. It was found that axial near-field intensity variations were smoothed compared with those of a uniform velocity distribution. Schafer and Lewin (1989) used the angular spectrum method (ASM) to predict and analyze particle velocity on the surface of a circular planar transducer and found a nonuniform velocity magnitude pattern. These authors used an impedance relationship to calculate the surface particle velocity, however Mair *et al.* (1987) showed that there were significant differences between the relative pressure squared and intensity profiles in the near field, especially close to the surface of the acoustic source.

The ASM uses pressure measurements in a certain plane to backward propagate the field to the source to determine the velocity distribution, and to forward propagate the field to predict the field beyond the measurement plane (Stepanishen and Benjamin, 1982). This work was later extended to axisymmetric transducers by Mederios and Stepanishen (1984) who reported nonuniform distributions of magnitude of velocity for a circular source. More recently, the ASM was extended to include nonlinear effects (Vecchio and Lewin, 1994) and to predict acoustic wave propagation through a layered medium (Vecchio *et al.*, 1994). However, this method may introduce errors imposed by the limits on the spatial and temporal sampling rates. Another technique that can be used to predict acoustic fields using forward and backward propagation is the superposition of Gaussian beams method (SGB) (Wen and Breazeale, 1988; Zhou *et al.*, 1996). However, this technique is suitable only for the analysis of an axially symmetric field, and large discrepancies may still exist in the near-field region.

The purpose of this article is to introduce a method that uses experimental measurements at a 2-D plane to accurately predict the entire acoustic field. This method uses the ultrasound phased array concept reported by Ebbini and Cain (1989). In this approach the single-source ultrasound transducer is treated as a phased array that is composed of many

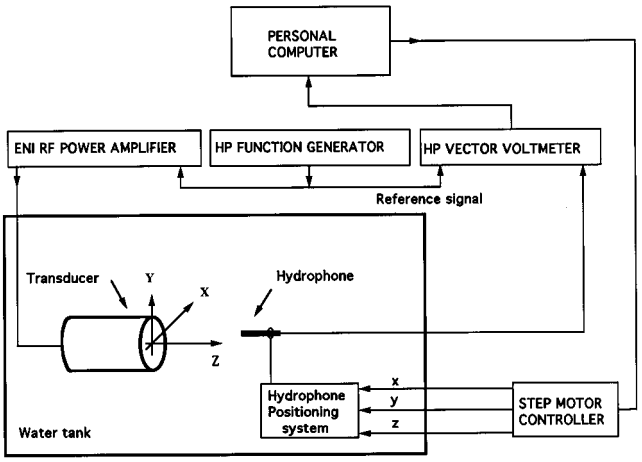


FIG. 1. Schematic of the overall measurement system and of the coordinate system used in this study.

small elements. The ultimate goal is to find the amplitude and phase for each phased array element that produces an acoustic field similar to the actual field generated by the original transducer. In other words, the goal is to determine an equivalent phased array model for the original single-source transducer. To accomplish this goal, a measured acoustic pressure field over a plane parallel to the face of the transducer in the extreme nearfield was used to calculate the amplitude and phase of the excitation source for an equivalent phased array using an inverse technique. Two circular and two square planar PZT transducers with operating frequencies in the range of 0.5–1.5 MHz were selected for this study. It is demonstrated that the equivalent phased array method (EPAM) provides good agreement between predicted acoustic field distributions and experimental results for planes at and beyond the measurement plane used for the calculation of the excitation source of the equivalent phased array.

I. METHODS

A. Ultrasound field calculation

The complex acoustic pressure field (p) due to a planar transducer surrounded by an infinite plane rigid baffle is given by the Rayleigh–Sommerfeld diffraction integral,

$$p(x,y,z)=\frac{j\rho ck}{2\pi}\int_s u\,\frac{e^{-jkr}}{r}\,ds, \tag{1}$$

where $j=\sqrt{-1}$, ρ is the density of the medium, c is the speed of sound in the medium, k is the real wave number, u is the particle velocity on the transducer surface, r is the distance from a point on the surface of the transducer to the field point of interest, and s is the area of the whole transducer. The coordinate system used in this study is shown in Fig. 1.

To use the phased array concept, the whole transducer face is divided into N elements, then Eq. (1) can be written in the following form:

$$p(x,y,z)=\frac{j\rho ck}{2\pi}\sum_{n=1}^{N} u_n\int_{s_n}\frac{e^{-jkr_n}}{r_n}\,ds_n$$

$$=\sum_{n=1}^{N} u_n\left(\frac{j\rho ck}{2\pi}\int_{s_n}\frac{e^{-jkr_n}}{r_n}\,ds_n\right), \tag{2}$$

where u_n is the complex excitation source of the nth element, r_n is the distance from a point on the nth element to the field point of interest, and s_n is the area of the nth element. The integral in Eq. (2) is evaluated using Huyghen's principle and summing the contributions from incremental areas composing an array element surface (Zemanek, 1971; Ochelteree and Frizzell, 1989). For the pressure at M field points (p_m, $m=1,...,M$) Eq. (2) can be written in matrix form:

$$\mathbf{p}=\mathbf{Hu}, \tag{3}$$

where

$$\mathbf{P}=\begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_M \end{bmatrix},\quad \mathbf{H}=\begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N} \\ h_{21} & h_{22} & \cdots & h_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M1} & h_{M2} & \cdots & h_{MN} \end{bmatrix},$$

$$\mathbf{u}=\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}.$$

Each entry of matrix **H** is calculated from the expression inside of the parentheses in Eq. (2). The physical meaning of h_{mn} is the complex pressure field at point m due to the nth transducer element whose particle velocity amplitude and phase are equal to 1 and 0, respectively. Equation (2) was used for the acoustic field calculations that were performed on a personal computer (Pentium 200 MHz, 128 Mb RAM, Micron Electronics, Inc., Nampa, ID) using FORTRAN (FORTRAN Power Station, version 4.0, Microsoft Corporation, Redmond, WA).

B. Excitation source calculation

Equation (3) can be used to calculate the excitation source for the equivalent phased array. The excitation source **u** can be obtained from

$$\mathbf{u}=\mathbf{H}^{\dagger}\mathbf{p}, \tag{4}$$

where matrix $\mathbf{H}^{\dagger}$ is the pseudoinverse of **H**, which was obtained by singular value decomposition (SVD) (Golub and Kahan, 1965).

The matrix **p** used to estimate **u** using Eq. (4) was obtained from the acoustic pressure (amplitude and phase) measurements in a water coupling medium for a given transducer. These measurements were taken in the extreme near field at a plane parallel to the face of the transducer. The area of the measurement plane was large enough so that it contained the majority of the propagation energy, but small enough to minimize extraneous noise. Another important requirement was that the number of measurement points in the measurement plane had to be larger than the number of

TABLE I. Planar ultrasound transducers used in the experiments and simulations.

Transducer	Shape	Size (mm)	Frequency (MHz)
T1	Circular	19.0 (diameter)	0.5
T2	Square	12.4 (side length)	1.0
T3	Circular	19.0 (diameter)	1.0
T4	Square	12.7 (side length)	1.5

equivalent phased array elements in order to determine the excitation source using Eq. (4). It is important to notice that under this requirement there can be more than one vector **u** that satisfies Eq. (3) for a given **p** and **H**. However, the vector **u** given by Eq. (4) is unique. In other words, this is an overdetermined problem for which the singular value decomposition technique provides the least squares error solution.

Another factor affecting the accuracy of the prediction is the number of elements composing the equivalent array. In order to predict well the experimental results using the EPAM, a suitable number of array elements must be used. In this study the number of array elements was increased until a good agreement between theoretical and experimental results was obtained. In general, half a wavelength was a reasonable choice for the array element size.

According to Eq. (3), the equivalent phased array with calculated excitation source should produce similar pressure fields at the measurement plane. The field distribution on this plane could be thought of as secondary sources. Since both measured and predicted field distributions at the measured plane are similar to each other, the measured and predicted fields beyond this plane should be similar to each other as well. In other words, the equivalent phased array with calculated excitation source is able to predict the acoustic field at and beyond the measurement plane. However, the field distribution between the face of the transducer and this plane may not be accurately predicted using the EPAM. This is because, as stated earlier, the calculation of **u** in Eq. (3) is overdetermined, and the acoustic wave propagates forward from the plane of secondary sources. The EPAM, as presented here, did not take into account nonlinear wave propagation and scattering effects.

II. EXPERIMENTS

A. Transducers

The planar transducers used in the experiments and simulations were two circular transducers (diameter of 19 mm, operating frequencies of 0.5 and 1 MHz) and two square transducers (side length of 12.4 mm and 12.7 mm, and operating frequencies of 1.0 and 1.5 MHz, respectively). All of the transducers (Table I) used in the experiments were air-backed PZT crystals (model EBL#4, Stevely Sensors, Inc., Hartford, CT) mounted on Plexiglass holders with silicon rubber cement used for anchoring and waterproofing. All the crystals had thin layers of silver electrodes on both sides. The solder point on the surface of the crystal was located at one corner for the square transducers and at the edge for the circular transducers.

B. Pressure field measurements

The relative pressure amplitude and phase (relative to input signal) were measured in a degased water tank by scanning a pre-amplified 0.6-mm-diam polyvinylidene fluoride (PVDF) needle hydrophone (model TNU001A, NTR Systems, Inc., Seattle, WA). The hydrophone's directional response is above -10 dB for frequencies between 0.5 and 1.5 MHz for incident angles within $\pm 60°$ relative to normal incidence according to the manufacturer's data. The transducer was driven by an rf power amplifier (ENI model 240L) with continuous-wave (cw) generated by a function generator (HP model 3325A). To minimize the multi-reflections (standing waves) inside the tank, rubber slabs were placed on the inside walls of the tank, and a reflector was used to deflect waves to an absorbing target. The input power was selected high enough to get a good signal-to-noise ratio and was kept low enough to avoid cavitation and nonlinear wave propagation. A vector voltmeter (HP model 8508A) was used to digitize both amplitudes and phases detected by the hydrophone. The overall process was controlled by a personal computer. A block diagram of the complete experimental system is shown in Fig. 1. The pressure fields needed for the excitation source calculation were measured at a plane 2 mm away from the face of a given transducer. This is the shortest distance possible to a transducer during the experiments due to the solder point. The other data sets were consistently obtained by scanning over planes parallel to the face of the transducer and along its axial plane.

III. RESULTS

A. Amplitude and phase of excitation source

The calculated results of amplitude and phase distributions of excitation sources are presented first. Figure 2 shows the surface plots of amplitudes and phases of excitation sources for the transducers given in Table I: (a, b) for T1, (c, d) for T2, (e, f) for T3, and (g, h) for T4. Figures on the left-hand side are amplitudes, and those on the right-hand side are phases. The size of transducer elements, size of measurement planes, and grid spacing used in the calculations and experiments are given in Table II. The calculated amplitude and phase distributions varied for each transducer. The phases fluctuated within the range $-180°$ and $+180°$.

B. Comparison of experimental and predicted acoustic fields

The above calculated excitation sources were used to predict the relative pressure fields for comparison with experimental results using Eq. (2). Figures 3, 5, 7, and 9 show the comparison between the predicted and measured relative pressure fields along the central axis and at an axial plane for transducers T1, T2, T3, and T4, respectively. Figures 4, 6, 8,

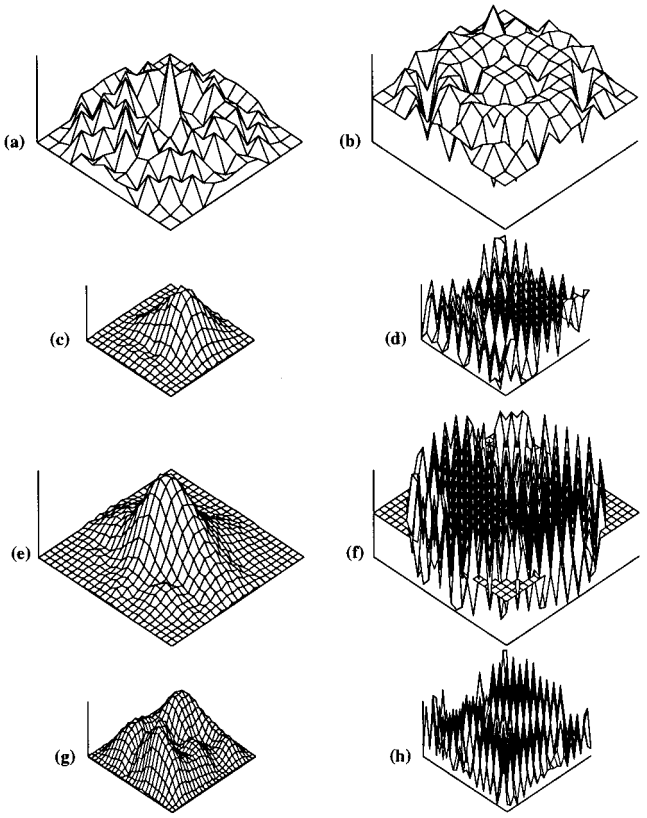


FIG. 2. Calculated excitation source distributions for the transducers given in Table I: (a, b) for T1, (c, d) for T2, (e, f) for T3, and (g, h) for T4. The figures on the left are magnitudes, and those on the right are phases.

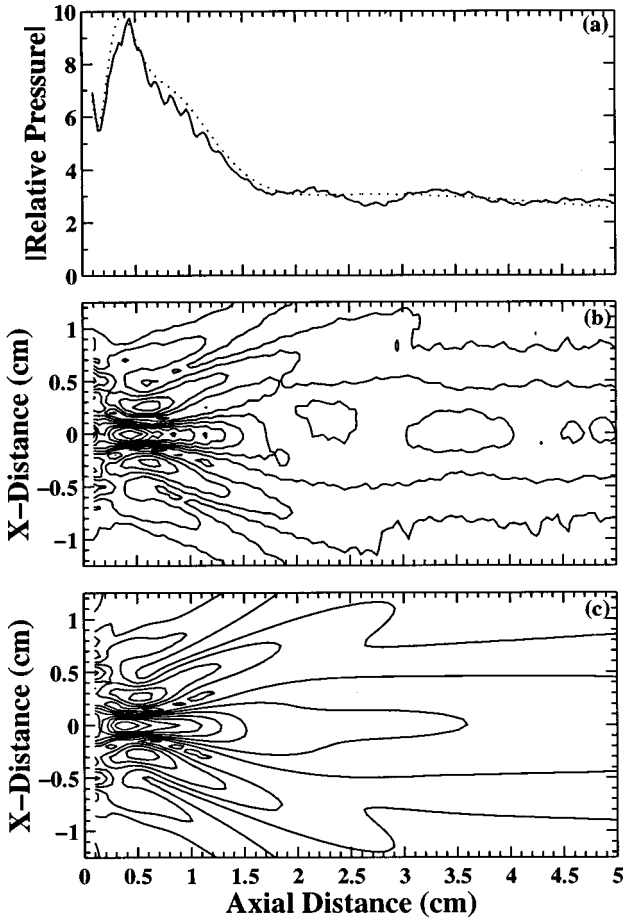


FIG. 3. Comparison between measured and predicted relative pressure fields for transducer T1: (a) measured (solid line) and predicted (dotted line) central axis distributions; (b) measured axial plane distribution; and (c) predicted axial plane distribution.

and 10 show the predicted and measured relative pressure fields at planes $z=2$ mm and $z=15$ mm for transducers T1, T2, T3, and T4, respectively.

For the circular transducer T1, as it can be seen in Fig. 3, the predicted fields matched well with the experimental results in major grating lobes and general distribution. One peak was observed on the central axis near $z=4$ mm. The global maximum (not shown here) was observed closest to the transducer during the experiments. The average pressure magnitude error along the central axis was about 5%. The predicted and measured relative pressure fields at $z=2$ mm and $z=15$ mm planes were almost identical (Fig. 4). For the square transducer T2, there was good agreement between predicted and measured fields (Fig. 5). The global maximum and nominal near-field–far-field boundary were observed on the axis near $z=2$ mm and $z=12.5$ mm, respectively. The average magnitude pressure error along the central axis was less than 5%. The predicted relative pressure fields at $z=2$ mm and $z=15$ mm planes were in good agreement with the measured results (Fig. 6), except at the edge of the transducer in the extreme near field ($z=2$ mm). For the circular T3, the location of major grating lobes and overall distribution were accurately predicted (Fig. 7), except there was

TABLE II. The size of transducer elements, number of phased array elements, the size of measurement planes, the grid spacing, and number of data points used in the experiments and excitation sources calculations for the transducers in Table I.

Transducer	Size of transducer element (mm^2)	Number of phased array elements	Size of measurement plane (mm^2)	Grid spacing (mm)	Number of data points
T1	1.30×1.30	137	20×20	0.50	1413[a]
T2	0.73×0.73	289	14×14	0.50	841
T3	0.76×0.76	489	20×20	0.50	1413[a]
T4	0.49×0.49	676	14×14	0.25	3249

[a]To minimize the extraneous noise, the measured data points outside of 20-mm-diam circle were not used in the excitation source calculations.

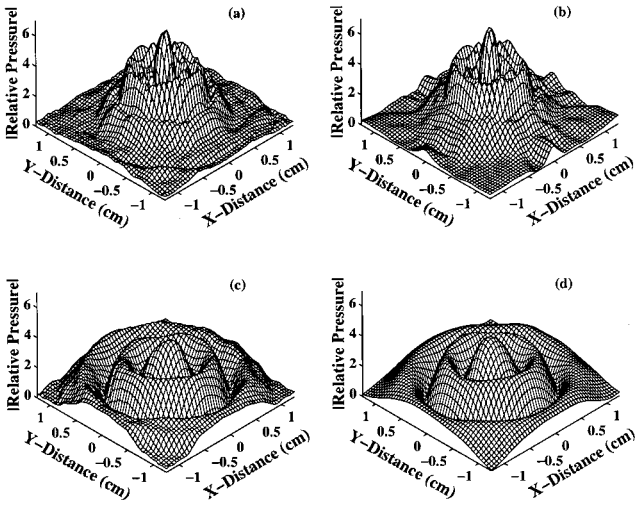


FIG. 4. Comparison between measured and predicted relative pressure fields at planes parallel to transducer T1: (a) plane measured at $z=2$ mm; (b) plane predicted at $z=2$ mm; (c) plane measured at $z=15$ mm; and (d) plane predicted at $z=15$ mm.

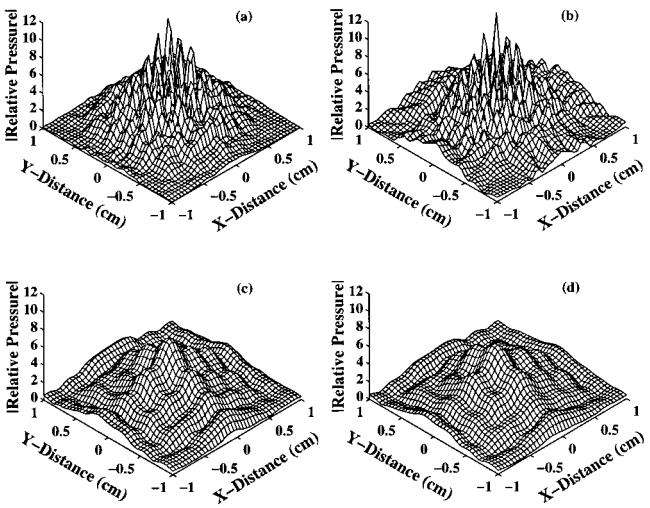


FIG. 6. Comparison between measured and predicted relative pressure fields at planes parallel to transducer T2: (a) plane measured at $z=2$ mm; (b) plane predicted at $z=2$ mm; (c) plane measured at $z=15$ mm; and (d) plane predicted at $z=15$ mm.

about 1-mm shift of the axial distribution. Two almost equal maxima were observed on the central axis near $z=2$ mm and $z=7$ mm. The average pressure magnitude error along the central axis was less than 15%. There was good agreement between the predicted and measured relative pressure fields

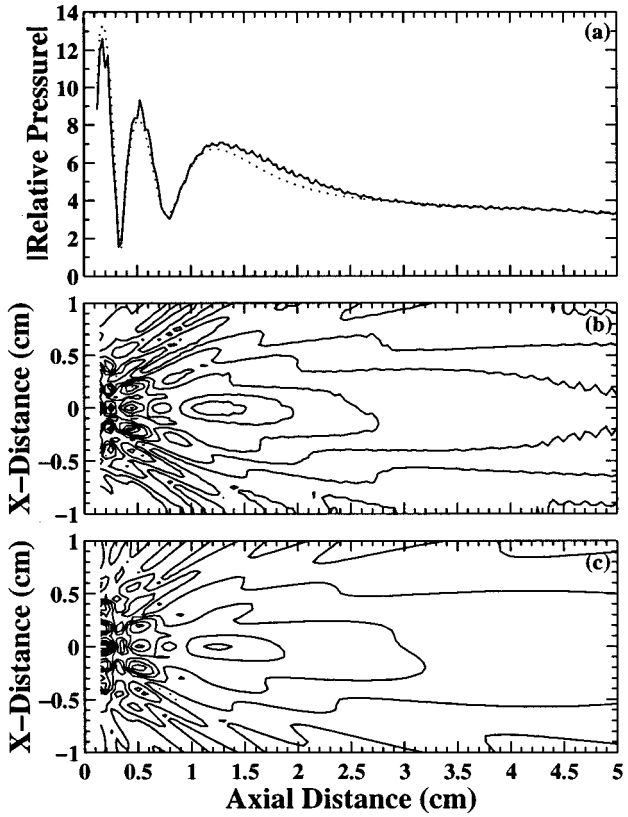


FIG. 5. Comparison between measured and predicted relative pressure fields for transducer T2: (a) measured (solid line) and predicted (dotted line) central axis distributions; (b) measured axial plane distribution; and (c) predicted axial plane distribution.

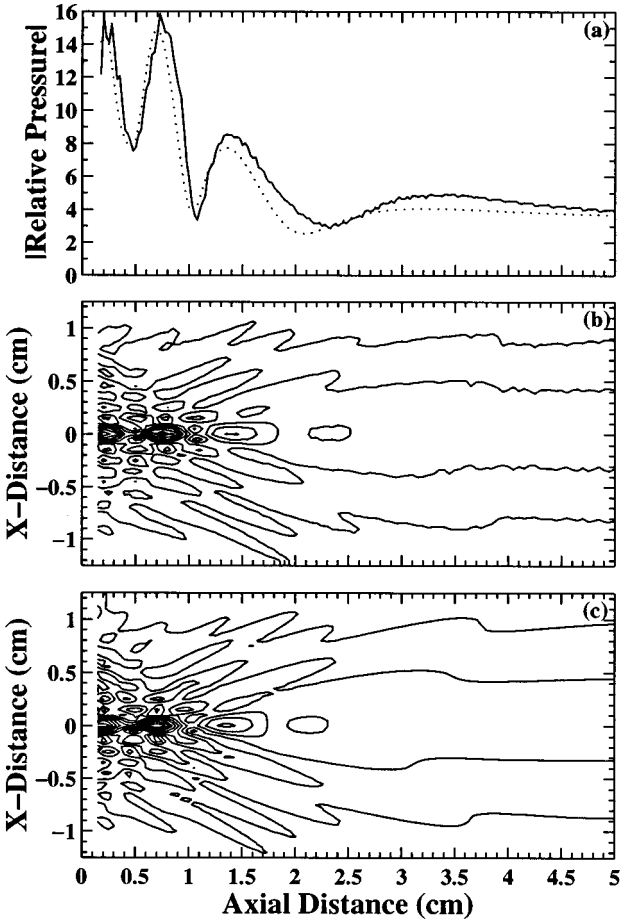


FIG. 7. Comparison between measured and predicted relative pressure fields for transducer T3: (a) measured (solid line) and predicted (dotted line) central axis distributions; (b) measured axial plane distribution; and (c) predicted axial plane distribution.

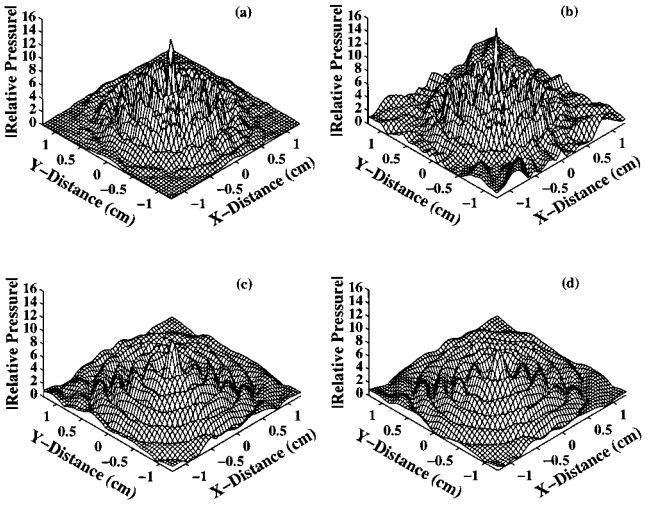


FIG. 8. Comparison between measured and predicted relative pressure fields at planes parallel to transducer T3: (a) plane measured at $z=2$ mm; (b) plane predicted at $z=2$ mm; (c) plane measured at $z=15$ mm plane; and (d) predicted at $z=15$ mm.

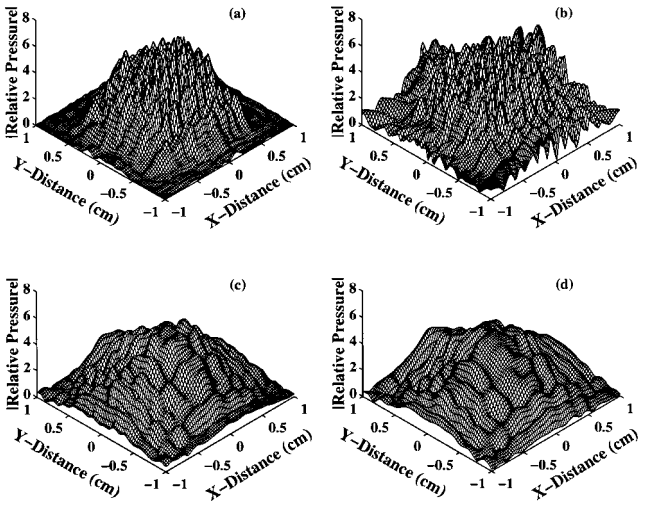


FIG. 10. Comparison between measured and predicted relative pressure fields at planes parallel to transducer T4: (a) plane measured at $z=2$ mm; (b) plane predicted at $z=2$ mm; (c) plane measured at $z=15$ mm; and (d) plane predicted at $z=15$ mm.

at $z=2$ mm and $z=15$ mm planes, except in the extreme near-field region ($z=2$ mm) around edge of the transducer (Fig. 8). Finally, for the square transducer T4, again, the location of major grating lobes and overall distribution were successfully predicted (Fig. 9). Two almost equal maxima were observed on the central axis near $z=1$ mm and $z=5$ mm. The average pressure magnitude error along the central axis was less than 5%. The predicted relative pressure field at the $z=15$ mm plane was similar to the measured field. High pressure values were predicted near the edge of T4 at the $z=2$ mm plane relative to the measured results (Fig. 10). This may be partially due to the size of hydrophone (0.6 mm), which is close to the wavelength (~ 1 mm) in water at this frequency.

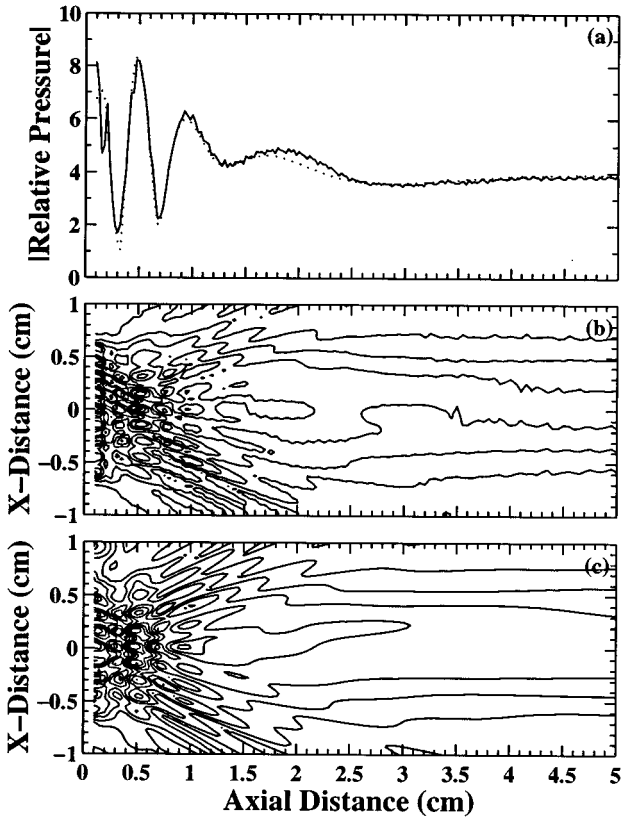


FIG. 9. Comparison between measured and predicted relative pressure fields for transducer T4: (a) measured (solid line) and predicted (dotted line) central axis distributions; (b) measured axial plane distribution; and (c) predicted axial plane distribution.

IV. DISCUSSION

From the results presented above, it can be seen that there was good agreement between the experimentally measured acoustic fields and those predicted by the EPAM. The results obtained by the EPAM accurately predicted the locations of major grating lobes and the overall field distribution. It is important to note that the calculated excitation source is not the same as the real physical particle velocity on the surface of the transducer. This is the excitation source of an equivalent phased array. Using the calculated excitation source to drive the equivalent phased array produces an acoustic field similar to the experimentally measured single-source field. The excitation source is not only dependent on the choice of the measurement plane, but also depends on the mathematical technique used in the calculations (Ebbini and Cain, 1989, 1991). The above results also demonstrate that a phased array with a sufficient number of elements has the capability to produce specified fields.

When the present method is used to predict the acoustic fields, it is critical to obtain the pressure matrix **P** properly. The EPAM is sensitive to the input pressure field, and therefore, this must be accurately measured at a selected plane. This plane should be parallel to the planar source and its distance from the source should be accurately measured. Different pressure measurements over the same plane may produce slightly different excitation source distributions due to

experimental variations. Any small shift or tilt of the hydrophone with respect to the source during the measurements will be reflected in the excitation source calculations and will degrade the predicting power of the method. In other words, the more accurate and noise free the pressure measurement is, the closer the predicted fields will be to the actual field generated by the single source. For example, if the measurement plane for transducer T4 (a worst case scenario in this study) is shifted by 0.25 mm, the predicted field would also be shifted approximately by the same distance; however, it will also have up to 25% error in the maximum pressure magnitude with respect to the nonshifted field. Obviously, the higher the frequency used, the more nearly accurate the measurements must be. Also, it is recommended that the measurement plane for the calculation of the excitation source should be as close as possible to the transducer because the EPAM will not accurately predict the field between the face of the transducer and the measurement plane.

Some advantages of this method are that it is relatively easy to understand and to use, and that the predicted fields directly reflect the magnitude of the measured results. The EPAM can also be used to predict the acoustic field for a transducer that has an irregular vibration on its face. For example, the effect of a solder point on the face of a transducer is automatically taken into account with the EPAM. In comparison to the ASM and SGB, a possible disadvantage of the EPAM is the requirement of large computer memory and a fast processor as the number of phased array elements and experimental data points increase (i.e., as the single-source size increases). However, for small, low-frequency transducers as those used in this study, the calculations were successfully performed on a personal computer (200-MHz processor). The longest computational time to get the pseudoinverse of the **H** matrix was approximately 10 h for T4.

The difference between measured and predicted acoustic fields may be mostly due to measurement errors, the resolution and accuracy of the scanning system, and the active size and directional response of the hydrophone. The maximum size (diameter) of the hydrophone should be at least less than half a wavelength in order to minimize the measurement errors due to phase cancellations and pressure averaging over the face of the hydrophone. For instance, our 0.6-mm-diam hydrophone was probably too large for measurements in the extreme near field at 1.5 MHz (transducer T4), for which the wavelength in water was about 1 mm. This could be the reason that high fields were predicted at the edge of this transducer at the extreme near field [Fig. 10(b)].

Another important factor that determines the predicting ability of the EPAM is the number of measured field points used in the computation of **u**. The number of measured field points in the measurement plane should be at least the number of phased array elements to get accurate results (Fan and Hynynen, 1996). This requirement is due to the overdetermined nature of the inverse technique. For instance, the number of measured data points used in the excitation source calculations for the transducers T2 and T3 were approximately three times the number of phased array elements (see Table II), thus the results were very satisfactory.

Although in this study the EPAM was applied to continuous-wave ultrasonic sources, theoretically the method should also be applicable to pulsed sources as long as the same data acquisition technique is used throughout all the measurements.

V. CONCLUSIONS

The acoustic field calculation for a single planar source using the EPAM has been presented. It was shown that the EPAM accurately predicted acoustic field distributions when compared to the experimental results. It is concluded that the acoustic field of a single-source ultrasound transducer can be successfully modeled as the field from a phased array with experimentally estimated excitation source. The EPAM is a powerful method for characterization of ultrasonic fields with possible applications in diagnostic imaging, therapeutic ultrasound, and nondestructive testing. The application of this method to planar and focused transducers for ultrasound therapy is currently being investigated in our laboratory.

ACKNOWLEDGMENTS

This work was supported by a DHHS/NCI Grant No. R29-CA63121. The authors wish to thank Charles Hamontree for his friendship and assistance during the experimental setups.

Ebbini, E., and Cain, C. A. (**1991**). ''A spherical-section ultrasound phased array applicator for deep localized hyperthermia,'' IEEE Trans. Biomed. Eng. **38**, 634–643.

Ebbini, E., and Cain, C. A. (**1989**). ''Multiple-focus ultrasound phased-array pattern synthesis: Optimal driving-signal distributions for hyperthermia,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **36**, 540–548.

Fan, X., and Hynynen, K. (**1996**). ''A study of various parameters of spherically curved phased arrays for noninvasive ultrasound surgery,'' Phys. Med. Biol. **41**, 591–608.

Freedman, A. (**1960**). ''Sound field of a rectangular piston,'' J. Acoust. Soc. Am. **32**, 197–209.

Golub, G., and Kahan, W. (**1965**). ''Calculating the singular values and pseudo-inverse of a matrix,'' J. SIAM Numer. Anal. **2**, 205–224.

Greenspan, M. (**1979**). ''Piston radiator: Some extensions of the theory,'' J. Acoust. Soc. Am. **65**, 608–621.

Hutchins, D. A., Mair, H. D., and Taylor, R. G. (**1987**). ''Transient pressure fields of PVDF transducers,'' J. Acoust. Soc. Am. **82**, 183–192.

Hutchins, D. A., Mair, H. D., Puhach, P. A., and Osei, A. J. (**1986**). ''Continuous-wave pressure fields of ultrasonic transducers,'' J. Acoust. Soc. Am. **80**, 1–12.

Lockwood, J. C., and Willette, J. G. (**1973**). ''High-speed method for computing the exact solution for the pressure variations in the near field of a baffled piston,'' J. Acoust. Soc. Am. **53**, 735–741.

Mair, H. D., Hutchins, D. A., and Puhach, P. A. (**1987**). ''Intensity fields of continuous-wave axisymmetric transducers,'' J. Acoust. Soc. Am. **81**, 328–334.

Marini, J., and Rivenez, J. (**1974**). ''Acoustic fields from rectangular ultrasonic transducers for nondestructive testing and medical diagnosis,'' Ultrasonics **12**, 251–256.

Medeiros, A. F., and Stepanishen, P. R. (**1984**). ''The forward and backward projection of acoustic fields from axisymmetric ultrasonic radiators using impulse response and Hankel transform techniques,'' J. Acoust. Soc. Am. **75**, 1732–1740.

Ocheltree, K. B., and Frizzell, L. A. (**1989**). ''Sound field calculation for rectangular sources,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **36**, 242–248.

Schafer, M. E., and Lewin, P. A. (**1989**). ''Transducer characterization using the angular spectrum method,'' J. Acoust. Soc. Am. **85**, 2202–2214.

Stepanishen, P. R. (**1971**). ''The time-dependent force and radiation impedance on a piston in a rigid infinite planar baffle,'' J. Acoust. Soc. Am. **49**, 841–849.
Stepanishen, P. R., and Benjamin, K. C. (**1982**). ''Forward and backward projection of acoustic fields using FFT methods,'' J. Acoust. Soc. Am. **71**, 803–812.
Vecchio, C. J., and Lewin, P. A. (**1994**). ''Finite amplitude acoustic propagation modeling using the extended angular spectrum method,'' J. Acoust. Soc. Am. **95**, 2399–2408.
Vecchio, C. J., Schafer, M. E., and Lewin, P. A. (**1994**). ''Prediction of ultrasonic field propagation through layered media using the extended angular spectrum method,'' Ultrasound Med. Biol. **20**, 611–622.
Wen, J. J., and Breazeale, M. A. (**1988**). ''A diffraction beam field expressed as the superposition of Gaussian beams,'' J. Acoust. Soc. Am. **83**, 1752–1756.
Weyns, A. (**1980**). ''Radiation field calculations of pulsed ultrasonic transducers. Part 1: Planar circular, square and annular transducers,'' Ultrasonics **18**, 183–188.
Williams, E. G., and Maynard, J. D. (**1982**). ''Numerical evaluation of the Rayleigh integral for planar radiators using the FFT,'' J. Acoust. Soc. Am. **72**, 2020–2030.
Zemanek, J. (**1971**). ''Beam behavior within the nearfield of a vibrating piston,'' J. Acoust. Soc. Am. **49**, 181–191.
Zhou, D., Peirlinckx, L., Lumori, M. L. D., and Van Biesen, L. (**1996**). ''Parametric modeling and estimation of ultrasonic fields using a system identification technique,'' J. Acoust. Soc. Am. **99**, 1438–1445.

Laser ultrasonic chirp sources for low damage and high detectability without loss of temporal resolution

Todd W. Murray
Department of Material Science and Engineering, The Johns Hopkins University, Baltimore, Maryland 21218

Kevin C. Baldwin
Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, Maryland 21218

James W. Wagner
Department of Material Science and Engineering, The Johns Hopkins University, Baltimore, Maryland 21218

(Received 3 October 1996; accepted for publication 23 July 1997)

Linear frequency-modulated (chirped) acoustic signals have been generated using a pulsed laser spatially modulated by an absorption mask at the surface of a test material. By distributing the laser energy over an area, instead of focusing it to a point or line source, the peak power density of the laser source can be kept below the damage threshold of the material. The corresponding chirped ultrasonic surface wave packet produced by the source, although extended in time, is detected and processed using a matched filtering technique which compresses the packet into a pulse, thus preserving temporal resolution for accurate time-of-flight measurements. Matched filter processing of the chirped wave packet has been compared with the same processing applied to a narrow-band tone burst wave packet. Processing of the chirped signal permits easy separation of overlapped return echoes which could not be resolved when narrow-band signals were used. Finally, by compressing the energy within a chirped signal to a single detection spike, an apparent 15-fold enhancement in signal-to-noise ratio is observed. [S0001-4966(97)01611-1]

PACS numbers: 43.38.Zp, 43.60.Qv [SLE]

INTRODUCTION

Signal detectability is arguably the key parameter to be optimized when designing a laser-based system for remote generation and detection of ultrasonic signals. The dependence of signal-to-noise ratio upon signal amplitude, signal bandwidth, and laser brightness (both for generating and detecting sound) is well understood.[1–5] Increasing the ultrasonic signal amplitude directly improves signal-to-noise ratio and, correspondingly, the detectability of a laser ultrasonic system. There are limits, however, to increasing the source laser power as a means of generating large ultrasonic amplitudes. In some configurations and for particular materials systems, it may not be possible to achieve a satisfactory signal-to-noise ratio at laser power densities below the damage threshold for the material being tested. As an alternative to single pulse excitation, phased array systems have been proposed and demonstrated which can enhance surface and bulk wave amplitudes to acceptably high levels without requiring any individual element of the array source to exceed the material ablation threshold.[6–9] Alternatively, periodic spatial array[10–14] and temporal modulation of a single source[15–16] have been demonstrated to produce tone burst acoustic signals to which a receiving detector can be tuned, thus providing enhanced immunity to broadband noise. Among these three methods for modifying the laser source for enhanced detectability, the spatial array methods for generation of narrow-band, tone burst ultrasound are the simplest and least expensive to implement. The disadvantage of these narrow-band methods, however, is that the temporal extent of the wave packet generated by such sources can lead to uncertainty in time-of-flight measurements important for many practical laser inspection and measurement applications.

A new laser ultrasonic generation technique has been investigated which uses a spatial mask in a manner similar to that which may be employed for generation of narrow-band, tone burst ultrasound. Instead of periodic spatial modulation at the surface of a test object, however, the mask employed for this technique is linearly frequency modulated as a function of position and correspondingly generates ultrasonic wave packets which are linearly frequency modulated (FM) as a function of time (chirped). Although the chirped mask launches an ultrasonic wave packet which is extended in time, not unlike an ultrasonic tone burst, the nature of the chirped signal is such that once detected, it can be processed by a matched filter algorithm which generates a single narrow spike corresponding to the time of arrival of the wave packet.

Pulse compression techniques were originally developed to overcome system power limitations in radar applications.[17–20] These same concepts have been used in the field of ultrasonic NDT. Conventional contact piezoelectric systems using pulse compression methods such as linear FM,[21] random noise,[22] and Barker and Golay codes[23,24] have

been demonstrated. Pulse compression techniques have also been successfully applied to electromagnetic-acoustic transducer (EMAT) systems.[25] More recently, ultrasonic pulse compression has been applied to medical imaging systems.[26,27] In all of the above applications, the use of pulse coding techniques provides a means of circumventing peak power limitations in the transducer (or in the system under inspection) while maintaining high temporal resolution and signal-to-noise ratio (SNR). In the present work, the use of linear FM pulse compression is extended to laser ultrasonic NDT where sensitivity issues are often a major obstacle limiting the use of such systems in practical NDT applications. Furthermore, the power limitation (ablation threshold) encountered in laser ultrasonic generation can be quite severe as it is dictated by the thermal and optical properties of the material under inspection. Thus the use of linear FM, as well as other pulse compression techniques, may provide a means of extending the range of laser ultrasonic systems to cases where low ablation threshold and poor sensitivity may have proven prohibitive.

I. THEORY

In filtering surface wave data, considerations include signal-to-noise ratio as well as extraction of useful information, such as arrival time and amplitude, from the signals arriving at the detection point. For many applications, it is not necessary that the shape of the waveform be preserved in the filtering process as long as amplitude and time-of-flight information can be extracted from the filtered signals. Taking this into account, the signal-to-noise ratio (SNR) that is to be maximized at some (arbitrary) instant in time (t_m) may be given as

$$(\mathrm{S/N}) = s_0^2(t_m)/n_0^2(t), \tag{1}$$

where $s_0(t_m)$ is the maximum signal amplitude and $n_0^2(t)$ is the mean squared noise amplitude. If the noise is white, as would be the case for a shot noise limited interferometric detection system, the SNR can be maximized using a matched filter.[28] The output of a matched filter is obtained by cross correlating the received signal with the generated signal. The maximum SNR is given as

$$(\mathrm{S/N})_{\max} = 2E/N_0, \tag{2}$$

where E is the total signal energy and $N_0/2$ is the noise spectral density.

In the laser ultrasonic generation case, the energy in the acoustic waveform is proportional to the incident laser energy.[3] When generating surface waves with a Q-switched laser pulse, the surface wave frequency content is determined primarily by the spot size of the generation pulse. A small spot size will give rise to a high-frequency surface wave, while a broad generation spot will give rise to a low-frequency surface wave. Thus to achieve high resolution and SNR, the energy in the incident pulse could be increased and focused to a sharp point or line. Unfortunately, the energy density of the input pulse is limited, though, by the ablation threshold of the material. One scheme to circumvent this difficulty is to distribute the generating pulse over a large area on the surface of a specimen. This allows for generation to remain in the thermoelastic regime while at the same time increasing the total energy in the acoustic signal, and thus increasing the SNR. In order to distribute the laser energy over a large area and still maintain good signal resolution, the energy can be modulated over the extent of the illuminated region.

Narrow-band surface waves have been obtained by illuminating the specimen with a sinusoidal intensity pattern with the generation laser.[10–13] When using narrow-band signals in a matched filter arrangement, the output of the matched filter consists of a sinusoidal signal of twice the duration of the input signal and modulated by a triangular envelope. The use of narrow-band generation has drawbacks in that ambiguities exist at multiples of the signal period (at the sidelobes) and difficulties arise when detecting multiple signals in cases where the correlation envelopes overlap. These problems can be overcome by using pulse coding techniques, examples of which include the use of amplitude or frequency modulated signals, or pseudo-random signals.[17–27] By imposing these pulse compression techniques, the effective bandwidth of the signal is increased and the width of the envelope of the signal, after matched filtering, is decreased.

Linear frequency-modulation (FM) signals have been used extensively in radar and ultrasonics applications. Such signals are especially useful in cases where a short duration, high-energy pulse cannot be generated or transmitted. The signal energy may then be distributed over a large period of time and then be compressed into a short pulse, after reception, with a matched-filter system. In the laser ultrasonics case, this consists of generating a FM surface wave pulse train with a spatially broad laser source and, after interferometric detection, compressing the wave train into a sharp spike through matched filtering. The general expression for a linear FM pulse is given by

$$W(t) = \cos[\omega_0 t + bt^2]. \tag{3}$$

This produces a chirp pulse in which the angular frequency is linearly swept from ω_0 to $\omega_0 + 2bT$, where T is the temporal pulse duration and $2b$ is the linear frequency sweep rate. The pulse width obtained after match filtering a linear FM pulse is inversely proportional to the FM pulse bandwidth, bT/π.

II. SYSTEM DESCRIPTION

The experimental configuration is shown in Fig. 1. To generate a linear FM acoustic wave, a transmission mask was placed near the surface of the specimen and was illuminated with an expanded beam from a pulsed Nd:YAG source. The mask was constructed by taking the desired linear FM waveform [Eq. (3)] as a template and then thresholding about the zero value. The resulting mask is shown in the lower left-hand corner of Fig. 1. The dimensions of the active area of the mask are 1.1 cm wide by 2.5 cm long. The incident laser pulse was expanded to over 1.1 cm in diameter in an attempt to achieve relatively uniform illumination over the entire width of the mask. The laser pulse width was 13 ns and the total energy transmitted by the mask was 30 mJ (approximately 25% of the laser energy incident on the mask). The

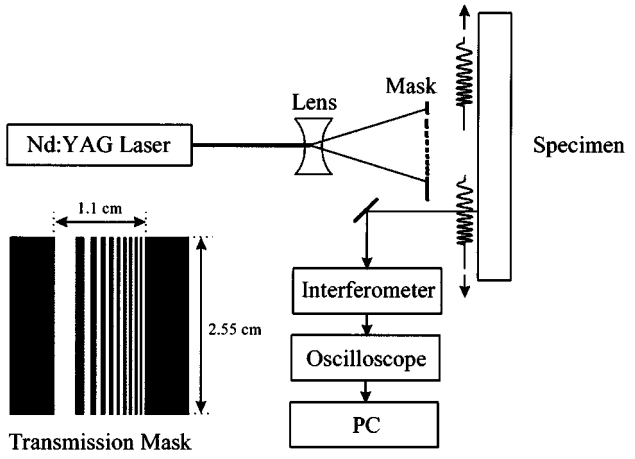


FIG. 1. Experimental setup for the generation of FM surface waves. The transmission mask is shown in the lower left-hand corner.

laser pulse energy was low enough to ensure that generation took place in the thermoelastic regime. All surface waves were generated on an aluminum specimen with multiple surface wave arrivals corresponding to reflections off of the edges. The acoustic waves were detected using a path-stabilized skewed Michelson interferometer. The waveforms were transferred to a personal computer for digital filtering.

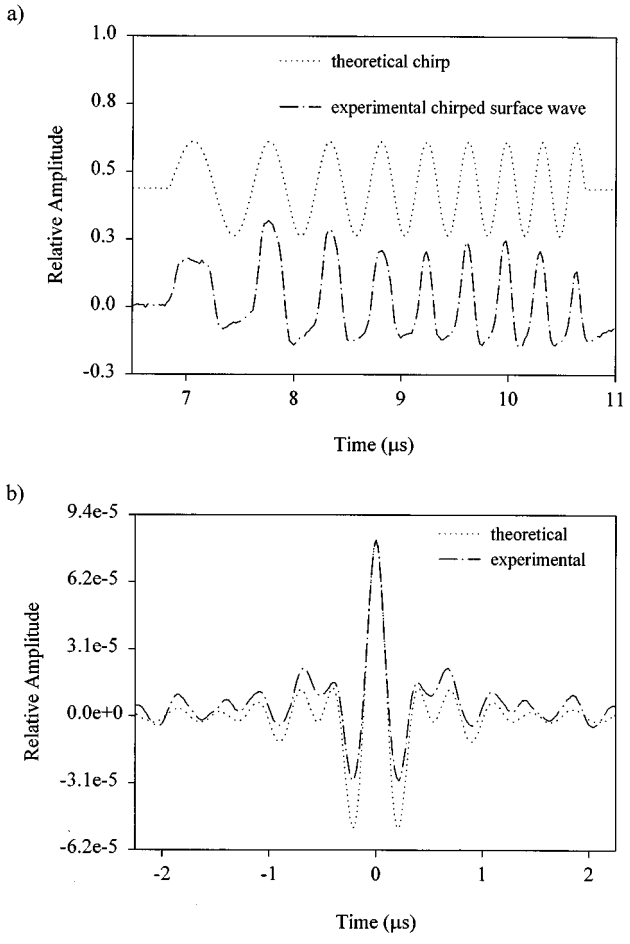


FIG. 2. Comparison of (a) theoretical and experimental chirped surface waves and (b) the compressed waveforms after autocorrelation.

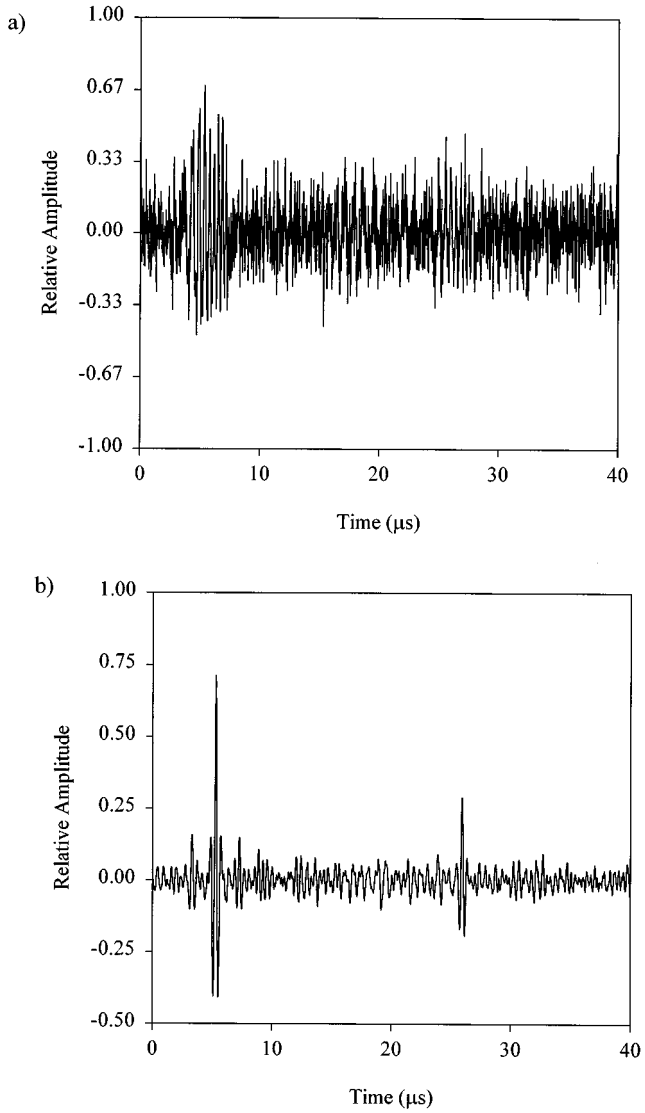


FIG. 3. (a) Single shot FM surface waveform and (b) the waveform after matched filtering showing a 15-fold increase in signal-to-noise ratio and compression of the wave packet.

III. RESULTS AND DISCUSSION

This technique launches acoustic waves that are very similar in form to the linear FM template, as is shown in Fig. 2(a) comparing the theoretical chirped signal to an experimental chirped surface wave which was signal averaged 100 times. The pulse compression obtained from the experimental and the theoretical FM waveform are in good agreement, as illustrated in Fig. 2(b). Note that, although the theoretical and experimental waveforms have some slight differences, the pulse compression achieved with the experimental signal is sufficient to prove that this techniques performs adequately. A typical single shot signal waveform using a linear FM mask is shown in Fig. 3(a). This signal was then processed using a matched filter based on the template linear FM waveform. The resulting output is shown in Fig. 3(b). An apparent 15-fold increase in SNR [based on Eq. (1)] is observed in the filtered waveform.

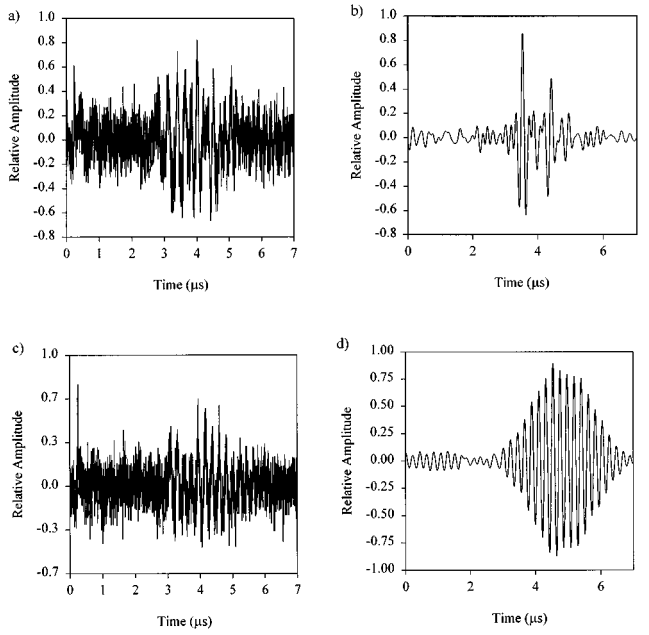


FIG. 4. Overlapping FM and narrow-band surface wave packets (a,c). Matched filtering allows for resolution of individual arrivals in the chirped case (b) but not in the narrow-band case (d).

A. Resolution enhancement

Analogous to the radar case, it becomes difficult to resolve distinct pulses in the output of a matched filter if the pulses are separated by less than full-width-half-maximum (FWHM) of their envelopes. In the narrow-band case, the envelope function is given by[18]

$$1-\frac{|t|}{T}, \quad |t|<T, \tag{4}$$

where T is the duration of the pulse train. In the linear FM case, the envelope is given by

$$\frac{\sin(bt[T-|t|])}{bt[T-|t|]}\left[1-\frac{|t|}{T}\right], \quad |t|<T. \tag{5}$$

As can be seen from the functional forms, the FWHM of the narrow-band case is T and the FWHM of the linear FM case is π/bT. The linear FM signal essentially ''compresses'' the envelope of the narrow-band signal by a factor of b/π.

The benefits of this compression effect are apparent in Fig. 4(a) shows a single shot ultrasonic signal containing two overlapping linear FM signals in noise. Figure 4(b) shows the result of using a linear FM matched filter. In this case, the compression of the envelopes allows the two signals to be easily resolved. Figure 4(c) shows a single shot ultrasonic signal composed of two overlapping narrow band signals in noise. Figure 4(d) shows the output of a matched filter based on a noise-free signal. Here, the width of the envelopes poses a problem in resolving the individual signals.

B. Directionality

Using the mask, a FM array of line sources generates a corresponding FM acoustic pulse train. Considering the surface wave directivity for a line source, large amplitude surface waves are generated perpendicular to illuminating lines.[29] These two FM pulse trains are phased reversed. Using a matched filter, the received signal can be correlated with a forward or reversed chirp signal. The acoustic signal launched to the left of the mask is phase reversed from that emitted to the right and thus will only correlate with a reversed chirp (see Fig. 1). This concept is illustrated in Fig. 5. Figure 5(a) shows the single shot data. Figure 5(b) shows the correlation with the forward chirp, with the peaks corresponding to surface waves originally launched to the right of the mask. Figure 5(c) shows the correlation with the reverse chirp, with the peaks corresponding to surface waves originally launched to the left of the mask. Note that in Fig. 5(c) there is some noise corresponding to a degree of correlation

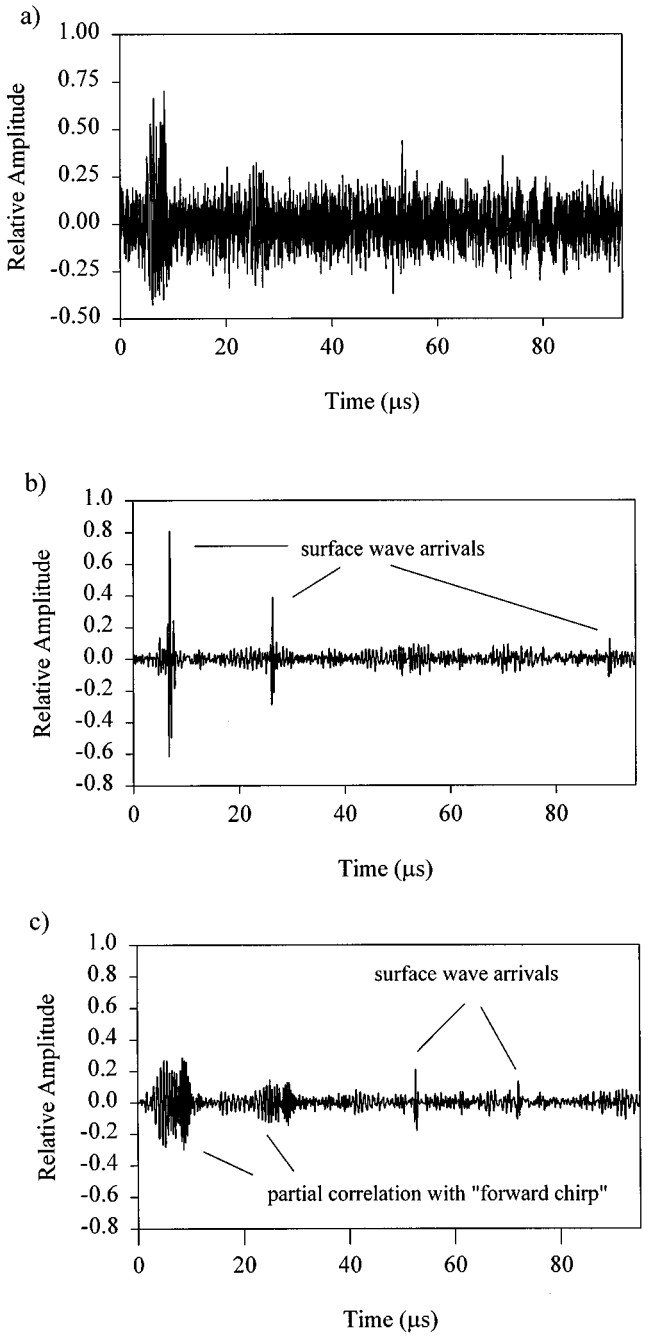


FIG. 5. (a) Single shot waveform on an aluminum block. Matched filtering with the ''forward chirp'' and ''reverse chirp'' (b), (c).

between the reverse chirp and the large amplitude surface wave pulse train originally launched to the right of the mask, but the broad nature of this noise (it is not compressed) allows it to be distinguished from the other arrivals. The asymmetry of the chirp pulse can prove useful when, for example, locating a crack when several arrivals are present.

IV. CONCLUSIONS

A new technique has been presented which allows for greater temporal resolution and enhanced detectability of laser generated surface waves. The laser energy is spread over an area of the surface allowing for more laser energy to be used in the generation of acoustic waves while remaining in the thermoelastic regime. Spatial frequency modulation of the laser source over the illuminated region allows for the compression of the generated surface wave packet upon reception by matched filtering. Matched filtering gives the maximum SNR possible in the white noise case. The filtered FM surface wave signals showed a 15-fold increase in SNR over the unfiltered waveforms. The advantages of using a FM technique over a narrow-band generation process, such as increased ability to resolve overlapping signals, have been discussed and experimentally verified.

The transmission mask used to generate the FM pulse train was placed in close proximity to the specimen. In practical inspection systems, where remote generation of surface waves is required, other means of producing a FM intensity distribution on the surface of the component to be inspected may be utilized. The transmission mask may be imaged onto the surface of the specimen using a simple imaging system. A binary phase grating, similar to that previously used for narrow-band surface wave generation,[10] may also be employed for the same purpose.

[1] C. B. Scruby and L. E. Drain, *Laser Ultrasonics, Techniques and Applications* (Adam Hilger, Bristol, 1990).

[2] D. A. Hutchins, ''Ultrasonic Generation by pulsed lasers,'' in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1988), Vol. XVIII, pp. 21–143.

[3] R. J. Dewhurst, D. A. Hutchins, S. B. Palmer, and C. B. Scruby, ''Quantitative studies of thermally generated elastic waves in laser irradiated metals,'' J. Appl. Phys. **51**, 6210–6216 (1980).

[4] J. W. Wagner, ''Optical detection of ultrasound,'' in *Physical Acoustics*, edited by E. Pierce (Academic, New York, 1990), Vol. XIX, pp. 201–266.

[5] J. W. Wagner and J. B. Spicer, ''Theoretical noise-limited sensitivity of classical interferometry,'' J. Opt. Soc. Am. B **4**, 1316–1326 (1987).

[6] M.-H. Noroy, D. Royer, and M. Fink, ''The laser-generated phased array: Analysis and experiments,'' J. Acoust. Soc. Am. **94**, 1934–1943 (1993).

[7] T. W. Murray, J. B. Deaton, Jr., and J. W. Wagner, ''Experimental evaluation of enhanced generation of ultrasonic waves using an array of laser sources,'' Ultrasonics **34**, 69–77 (1996).

[8] J. S. Steckenrider, T. W. Murray, J. W. Wagner, and J. B. Deaton, Jr., ''Sensitivity enhancement in laser ultrasonics using a versatile laser array system,'' J. Acoust. Soc. Am. **97**, 273–279 (1995).

[9] Y. Yang, N. DeRidder, C. Ume, and J. Jarzynski, ''Noncontact optical fibre phased array generation of ultrasound for non-destructive evaluation of materials and processes,'' Ultrasonics **31**, 387–394 (1993).

[10] J. Huang, S. Krishnaswamy, and J. D. Achenbach, ''Laser generation of narrow-band surface waves,'' J. Acoust. Soc. Am. **92**, 2527–2531 (1992).

[11] S. Nagai and H. Nakano, ''Laser generation of antisymmetric lamb waves in thin plates,'' Ultrasonics **29**, 230–234 (1991).

[12] A. Harata, H. Nishimura, and T. Sawada, ''Laser-induced surface acoustic waves and photothermal surface gratings generated by crossing two pulsed laser beams,'' Appl. Phys. Lett. **57**, 132–134 (1990).

[13] H. Nishino, Y. Tsukahara, Y. Nagata, T. Koda, and K. Yamanaka, ''Excitation of high frequency surface acoustic waves by phase velocity scanning of a laser interference fringes,'' Appl. Phys. Lett. **62**, 2036–2038 (1993).

[14] A. D. W. McKie, J. W. Wagner, J. B. Spicer, and C. M. Penny, ''Laser generation of narrowband and directed ultrasound,'' Ultrasonics **27**, 323–330 (1989).

[15] J. W. Wagner, J. B. Deaton, Jr., and J. B. Spicer, ''Generation of ultrasound by repetitively *Q*-switching a pulsed Nd:YAG laser,'' Appl. Opt. **27**, 4696–4700 (1988).

[16] J. B. Deaton, Jr., A. D. W. McKie, J. B. Spicer, and J. W. Wagner, ''Generation of narrow band ultrasound with a long cavity mode-locked Nd:YAG laser,'' Appl. Phys. Lett. **56**, 2390–2392 (1990).

[17] C. E. Cook, ''Linear *f-m* Pulse Compression,'' in *Modern Radar*, edited by R. S. Berkowitz (Wiley, New York, 1965), pp. 216–241.

[18] H. Urkowitz, ''Ambiguity and resolution,'' in *Modern Radar*, edited by R. S. Berkowitz (Wiley, New York, 1965), pp. 197–213.

[19] A. W. Rihaczek, *Principles of High Resolution Radar* (McGraw-Hill, New York, 1969).

[20] C. E. Cook and M. Bernfeld, *Radar Signals* (Academic, New York, 1967).

[21] F. K. Lam and J. Szilard, ''Pulse compression techniques in ultrasonic non-destructive testing,'' Ultrasonics **14**, 111–114 (1976).

[22] E. S. Furgason, V. L. Newhouse, N. M. Bilgutay, and G. R. Cooper, ''Application of random signal correlation technique to ultrasonic flaw detection,'' Ultrasonics **13**, 11–17 (1975).

[23] W. H. Chen and D. L. Deng, ''Ultrasonic nondestructive testing using Barker code pulse compression techniques,'' Ultrasonics **26**, 23–26 (1988).

[24] B. B. Lee and E. S. Furgason, ''High-speed digital Golay code flaw detection system,'' Ultrasonics **21**, 153–161 (1983).

[25] H. M. Frost, ''Electromagnetic-ultrasound transducers: Principles, practice, and applications,'' in *Physical Acoustics*, Vol. 14, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1979), pp. 179–276.

[26] J. Y. Chapelon, ''Pseudo-random correlation imaging and systems characterization,'' in *Progress in Medical Imaging*, edited by V. L. Newhouse (Springer-Verlag, New York, 1988), Chap. 6.

[27] N. A. H. K. Rao, ''Investigation of a pulse compression technique for medical ultrasound: a simulation study,'' Med. Biol. Eng. Comput. **32**, 181–188 (1994).

[28] A. D. Whalen, *Detection of Signals in Noise* (Academic, Orlando, FL, 1971), pp. 167–175.

[29] A. M. Aindow, R. J. Dewhurst, and S. B. Palmer, ''Laser generation of directional surface acoustic wave pulses in metals,'' Opt. Commun. **42**, 116–120 (1982).

Multi-variable structural acoustic control with static compensation

Robert L. Clark
Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708-0300

David E. Cox
Guidance and Controls Branch, NASA Langley Research Center, Hampton, Virginia 23681-0001

(Received 17 August 1996; accepted for publication 29 July 1997)

Structural acoustic control of stochastic disturbance sources based on static compensation of multiple-input, multiple-output colocated transducer arrays is addressed in this work. Prior feedback controller designs for structural acoustic systems have been based on linear quadratic regulator theory or linear quadratic Gaussian theory, however, the need for a state estimator limits the practical implementation of such control systems on plants characterized by model uncertainty and high modal density. Results presented in this work serve to demonstrate that direct output-feedback control can be successfully employed to attenuate the structural acoustic response of a panel subjected to exogenous input disturbances. Colocated transducer pairs were implemented with a fully populated constant feedback gain matrix from sensor outputs to control inputs. The colocated transducer array serves to enhance robust stability, while the constant feedback gain matrix provides a practical means of implementation. The effects of instrumentation density, control effort penalty, and model order on predicted performance are considered.
[S0001-4966(97)03611-4]

PACS numbers: 43.40.Vn [PJR]

INTRODUCTION

The active control of interior noise on modern aircraft is directed at both periodic and stochastic noise sources. If an uncontrollable reference signal correlated with the disturbance source is available, adaptive feedforward control approaches are typically employed. However, for stochastic sources such as turbulent boundary layer (TBL) noise, such reference signals are not available. One must resort to feedback control strategies in an attempt to increase the transmission loss across external panels. The use of static compensators (constant gain feedback) for structural acoustic control is the focus of this work.

It is well-known that structural acoustic control differs from vibration control, specifically below coincidence of panel modes, since there is significant interaction between the modes in the radiated field (Fuller, 1988). Baumann *et al.* (1991, 1992) explored this frequency-dependent structural acoustic coupling, described by Borgiotti (1990) and Photiadis (1990), to design radiation filters. The radiation filters were then used to weight the modal velocities of the structure such that linear quadratic regulator (LQR) theory and linear quadratic Gaussian (LQG) theory could be used to design the control system. The LQR approach assumes that full information concerning the system states is available, and the LQG approach is based on the design of a state estimator. In practice, the realization of a control system based on an estimator is quite involved for plants with high modal density and augmented with radiation filters. However, the LQR approach does provide useful information concerning the maximum achievable performance for a given transducer array and will be used in this study for that purpose.

The purpose and contribution of this work is to detail the performance which can be achieved through the use of static (constant gain feedback) compensation on square (equal inputs and outputs) systems with colocated transducers. The dual Levine–Athans method (Toivonen, 1985; Makila and Toivonen, 1987) was implemented to design a static compensator which was optimal with respect to a frequency weighted cost function and chosen control effort penalty. The frequency-dependent weighting of the output was achieved by developing radiation filters as outlined by Baumann *et al.* (1991) and implementing them as detailed by Gupta (1980) to augment the system. The radiation filters were realized by spectral factorization and were used to weight the performance output. For the given frequency weighted cost functional, a static compensator was designed and compared to the LQR controller for various input–output configurations.

Results from this study are based on an analysis of a plate with pinned boundary conditions, bordered by an infinite baffle. The system model contained 216 states, 80 of which were due to the 40 structural modes included in the model and the remainder due to the radiation filters. Radiation filters were designed for the first ten modes of the panel and the potential effects of spillover are considered. Colocated, point force actuators and point velocity sensors were used to implement the control system with actuator arrays ranging from single-input, single-output (SISO) to five-input, five-output (5I5O). The effect of constraining control effort in the synthesis was also considered to determine appropriate performance limitations. Results from this study demonstrate

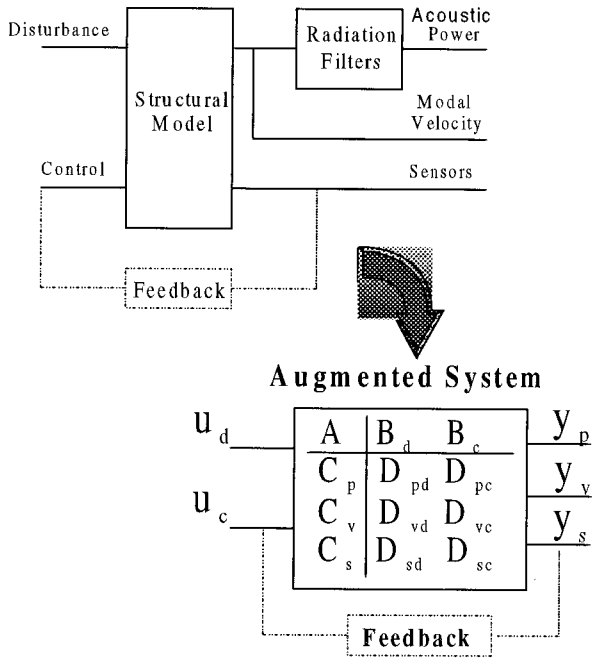


FIG. 1. Block diagram of structural acoustic system.

that multi-input, multi-output (MIMO) feedback control with static compensation provides a viable alternative to dynamic compensation. Advantages of static compensation are less complexity in the control system implementation and the ease of generating a dissipative controller which is robust to model uncertainty.

I. MODEL DEVELOPMENT

The block diagram in Fig. 1 shows how the different subsystem models are connected and used to form the model for control design and performance analysis. This augmented model has two major dynamic subsystems: the structural model and the radiation filters. The development of each of these systems is described in this section. Inputs to the system were taken to be point force disturbances and point force control actuators. The system's outputs were the radiated power, modal velocities, and point velocity sensors (colocated with the point force control actuators).

By selecting different input–output combinations from the augmented system different models could be used for control law design and analysis. For feedback design only the sensors were assumed available as signals for the control law, however the control law was optimized with respect to the radiated power outputs. Actual performance was evaluated by computing the rms power from a set of disturbance inputs to the radiated acoustic power. Modal velocity outputs were used to evaluate the extent of structural spillover caused by the control law. Spillover here refers to modes which were included in the structural model but considered negligible for the acoustic radiation filters, and therefore were not fed through to the performance output. The evaluation of these different models is discussed in Sec. IV.

A. The structural model

The approximation of a continuous, conservative, nongyroscopic system into a discrete system can be found in the reference by Meirovitch (1967). The objective is to develop an accurate low-frequency model of a reverberant plant with reasonable convergence of both the poles and zeros in independent transfer function loops as outlined by Clark (1995). Lagrange's equations of motion can be used to express the system response as follows:

$$\sum_{r=1}^{\infty} m_{nr}\ddot{q}_r(t)+\sum_{r=1}^{\infty} k_{nr}q_r(t)=Q_n(t), \quad n=1,2,\ldots, \tag{1}$$

where $q_r(t)$ is the response of the r-th generalized coordinate, m_{nr} is the mass coupling between the nth and rth generalized coordinate, and k_{nr} is the stiffness coupling between the nth and rth generalized coordinate. The displacement response is expressed in terms of a series expansion:

$$w(\mathbf{x},t)=\sum_{n=1}^{\infty} \phi_n(\mathbf{x})q_n(t), \tag{2}$$

where $\mathbf{x}$ is the spatial location of the response in general three-dimensional coordinates, $w(\mathbf{x},t)$ is the time-dependent displacement at position $\mathbf{x}$, and $\phi_n(\mathbf{x})$ is the n-th admissible function. In the event that the chosen admissible functions are actually eigenfunctions, the system of Eqs. (1) are decoupled.

Consider, for now, the special case in which the eigenfunctions are known. The response at the jth output can be expressed as follows:

$$y_j(\mathbf{x},t)=\sum_{n=1}^{\infty} q_n(t)\int_D \phi_n(\mathbf{x})S_e(\mathbf{x}_j)dD(\mathbf{x}), \tag{3}$$

where $S_e(\mathbf{x}_j)$ is the spatial aperture of the response to be measured at coordinate $\mathbf{x}_j$ (Burke and Hubbard, 1990). Likewise, the generalized force is expressed in terms of a series expansion of disturbance and control forces distributed force over the domain:

$$Q_n(t)=\sum_{i=1}^{N_d} u_i^d(t)\int_D S_f(\mathbf{x}_i)\phi_n(\mathbf{x})dD(\mathbf{x}) + \sum_{i=1}^{N_c} u_i^c(t)\int_D S_f(\mathbf{x}_i)\phi_n(\mathbf{x})dD(\mathbf{x}), \tag{4}$$

where $u_i^d(t)$ is the time-varying amplitude of the ith disturbance force, $u_i^c(t)$ is the time-varying amplitude of the ith control force, $S_f(\mathbf{x}_i)$ describes the spatial aperture associated with each input force, N_d is an integer describing the number of disturbances, and N_c is an integer describing the number of control forces. For a point force, $S_f(\mathbf{x}_p)=\delta(\mathbf{x}-\mathbf{x}_p)$, where $\mathbf{x}_p$ is the coordinate of the input force.

Once the type of structure and associated boundary conditions are identified, a state-variable model of the system can be formed. For the purpose of this study, a plate with pinned boundary conditions was used and the admissible functions chosen as

$$\phi_n(\mathbf{x})=\sin(n_x\pi x_x/L_x)\sin(n_y\pi x_y/L_y), \tag{5}$$

where x_i is the ith component of the coordinate vector $\mathbf{x}$, n_x and n_y are indices of the nth admissible function, L_x and L_y are the dimensions of the panel in the x and y directions respectively, and $\phi_n(\mathbf{x})$ is the nth admissible function which in this case also corresponds to the nth eigenfunction of the homogeneous plate in the absence of discontinuities. The state-variable model of the structure can be formed by defining the state vector

$$\mathbf{z}_{2N\times 1}=\begin{bmatrix}\mathbf{q}_{N\times 1}\\ \dot{\mathbf{q}}_{N\times 1}\end{bmatrix}, \tag{6}$$

and rewriting a finite sum approximation to Eq. (1) as a set of first-order matrix equations of order N. State-variable models were constructed for both the plate and radiation filters and combined to form the augmented system as illustrated in Fig. 1.

B. Radiation filters

The panel with pinned boundary conditions is assumed to reside within a baffle of infinite extent. The sound radiation from this structure can thus be predicted from Rayleigh's integral and as outlined by Lomas and Hayek (1977), Baumann *et al.* (1991), and Cunnefare (1991) the self- and mutual radiation efficiencies of the panel modes can be predicted. Below coincidence, the structural modes interact acoustically and this coupling can be used to control the sound power radiated from the structure. Above coincidence, the structural modes are orthogonal in the acoustic field and radiate independently with uniform radiation efficiency. For active structural acoustic control (ASAC), we choose to explore the bandwidth over which there is significant interaction between modes. Radiation filters were originally introduced by Baumann *et al.* (1991) for ASAC such that the dynamic response of the system could be augmented to incorporate additional states (radiation modes) which provide a measure of rms power in the LQR/LQG formulation. The augmentation of system dynamics in the formulation of a ''frequency shaped'' cost was originally proposed and detailed by Gupta (1980) and was extended to other exemplary applications.

For the infinite baffle, a frequency-dependent expression for the power radiated can be rendered by integrating an expression for the acoustic intensity over the hemisphere. Since the pressure can be obtained from Rayleigh's integral, the far-field acoustic intensity in the half-space is easily obtained. The form of the radiated power is:

$$\Pi=\frac{1}{\pi}\int_0^\infty \mathbf{V}^H(j\omega)\mathbf{M}(j\omega)\mathbf{V}(j\omega)d\omega, \tag{7}$$

where Π is the total radiated power, $\mathbf{V}(j\omega)$ is a matrix of the Fourier transform of modal velocities, and $\mathbf{M}(j\omega)$ represents the frequency-dependent interaction between structural modes required to compute the radiated power. The self- and mutual radiation efficiencies of the structural modes can be readily obtained from this matrix as outlined by Lomas and Hayek (1977) and more recently by Cunnefare (1991).

If the Laplace domain equivalent of the matrix $\mathbf{M}(j\omega)$ is expressed as $\mathbf{M}(s)$, then one can develop the matrix in terms of a filter and its parahermitian conjugate:

$$\mathbf{M}(s)=\mathbf{G}^T(-s)\mathbf{G}(s), \tag{8}$$

where $\mathbf{G}(s)$ is the radiation filter as detailed by Baumann *et al.* (1991). This filter can be obtained through spectral factorization as outlined by Francis (1986), and a state-variable model for the filter can be used to augment the system dynamics. The time-averaged sound power radiated from the structure when driven by a white noise excitation can thus be computed from the output of these augmented states.

The radiation filters were represented schematically in Fig. 1, and each filter serves to weight the modal velocities of the structure. For the work herein, the radiation filters were designed from the self- and mutual radiation efficiencies, providing a normalized measure of the rms power. This normalization only effects the choice of $\mathbf{Q}$ in the design optimization process. For comparison, the squared magnitude of several exemplary radiation filters are compared to respective plots of self- and mutual radiation efficiencies from which they were derived in Fig. 2. The squared magnitude of the filters are represented by bold lines while the computed radiation efficiency is represented by a thin line. The self- and mutual radiation efficiencies of the selected admissible functions are indicated in the parentheses next to the appropriate curve with the first two indices representing a particular admissible function and the second two indices representing the interacting function. As illustrated, the dynamics of the radiation filters accurately capture the frequency response of the self- and mutual radiation efficiencies.

II. CONTROLLER SYNTHESIS

For the purpose of comparison, two different types of control laws, state feedback and output gain feedback, were investigated. Both sought to minimize a linear quadratic cost function which balanced performance with control effort. State feedback methods are well-suited for this problem and optimal LQR gains provide an upper bound to obtainable performance. However, the same problem can be posed with constraints that the control action be simply a constant gain applied to the sensor signals. Although this type of control is far more restrictive than state feedback, the ease of implementation makes its consideration important.

A. State feedback control

The optimal control problem for the system,

$$\dot{\mathbf{x}}=\mathbf{A}\mathbf{x}+\mathbf{B}_c\mathbf{u}_c, \tag{9}$$

$$\mathbf{y}_p=\mathbf{C}_p\mathbf{x}, \tag{10}$$

is to find the input signal, $\mathbf{u}_c$, which minimizes the cost function

$$J=\int_0^\infty \mathbf{y}_p^T\mathbf{Q}\mathbf{y}_p+\mathbf{u}_c^T R\mathbf{u}_c\, dt, \tag{11}$$

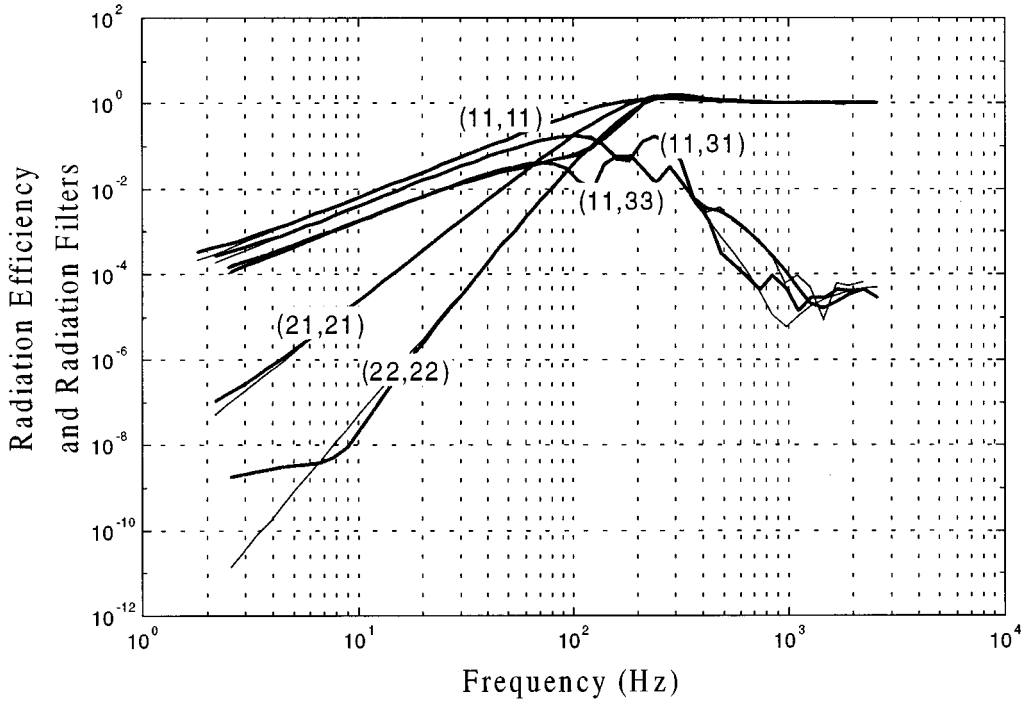


FIG. 2. Comparison of self- and mutual radiation efficiencies of structural modes with squared magnitude of radiation filters.

where $\mathbf{Q}$ is a semi-positive definite performance weighting, R is the positive definite control effort penalty, $\mathbf{y}_p$ is the performance output, and $\mathbf{u}_c$ is the actuator control signal, as shown in Fig. 1. The performance output is a weighted measure of the acoustic radiation and can be written as a linear function of the augmented states, i.e., $\mathbf{y}_p = \mathbf{C}_p\mathbf{x}$. Therefore Eq. (11) can be written,

$$J = \int_0^\infty \mathbf{x}^T\tilde{\mathbf{Q}}\mathbf{x} + \mathbf{u}^T R\mathbf{u}\, dt, \tag{12}$$

where $\tilde{\mathbf{Q}} = \mathbf{C}_p^T\mathbf{Q}\mathbf{C}_p$. For state feedback the input is of the form $\mathbf{u} = -\mathbf{K}\mathbf{x}$, and the minimization of Eq. (11) yields the well-known linear quadratic regulator (Kalman *et al.*, 1963; Anderson *et al.*, 1990). This approach is often used for generating an optimal linear control law, and its application to structural acoustic problems has been detailed elsewhere (Baumann *et al.*, 1991; Thomas and Nelson, 1995).

Implementation of a state feedback control law requires full knowledge of the state, which typically must be estimated from a limited set of measurements. When employing radiation filters, especially those derived from curve fitting, the augmented plant can become very high order. This makes practical implementation of a state estimator difficult. Nevertheless, the state feedback gains provide an upper bound for the performance of the simpler constant gain controllers. They are computed here for comparison and closed loop results are derived assuming perfect state information is available.

B. The dual Levine–Athans algorithm

The minimization of Eq. (11) can also be carried out under the assumption that $\mathbf{u} = \mathbf{G}\mathbf{y}_s = \mathbf{G}\mathbf{C}_s x$, where $\mathbf{y}_s$ is the output signal from available measurements. This implies that the control signal is generated as a constant gain matrix times the sensor signals. This choice of control input forms a parametric LQ problem, and several techniques of obtaining the optimal controller have been investigated (Makila and Toivonen, 1987). However, unlike the state feedback case there is still no guaranteed unique solution. The approach taken here is a variation on the method proposed by Levine and Athans (1970), and its development follows Toivonen (1985). It has the advantage of being globally convergent to a solution from a stabilizing initial guess, and the solution satisfies the necessary conditions for optimality.

Consider the cost function in Eq. (11), by letting $\mathbf{u} = \mathbf{G}\mathbf{C}_s\mathbf{x}$ we can write

$$J = \int_0^\infty \mathbf{x}^T(\tilde{\mathbf{Q}} + \mathbf{C}_s^T\mathbf{G}^T R\mathbf{G}\mathbf{C}_s)\mathbf{x}\, dt. \tag{13}$$

This cost function depends on initial conditions of the state. Given random initial conditions with $E[\mathbf{x}(0)\mathbf{x}(0)^T] = \mathbf{P}$, which is equivalent to minimizing the expected value of the cost from a persistent input, the cost can then be expressed using the trace operator as

$$J = \mathrm{tr}((\tilde{\mathbf{Q}} + \mathbf{C}_s^T\mathbf{G}^T R\mathbf{G}\mathbf{C}_s)\mathbf{L}) = \mathrm{tr}(\mathbf{K}\mathbf{P}), \tag{14}$$

where $\mathbf{K}$ and $\mathbf{L}$ satisfy the equations,

$$0 = \mathbf{K}\mathbf{A}_c + \mathbf{A}_c^T\mathbf{K} + \mathbf{C}_s^T\mathbf{G}^T R\mathbf{G}\mathbf{C}_s + \tilde{\mathbf{Q}}, \tag{15}$$

$$0 = \mathbf{L}\mathbf{A}_c + \mathbf{A}_c^T\mathbf{L} + \mathbf{P}, \tag{16}$$

with $\mathbf{A}_c = \mathbf{A} - \mathbf{B}\mathbf{G}\mathbf{C}_s$. The gain

$$\mathbf{G}_{\mathrm{opt}} = R^{-1}B^T\mathbf{K}\mathbf{L}\mathbf{C}_s^T(\mathbf{C}_s\mathbf{L}\mathbf{C}_s^T)^{-1} \tag{17}$$

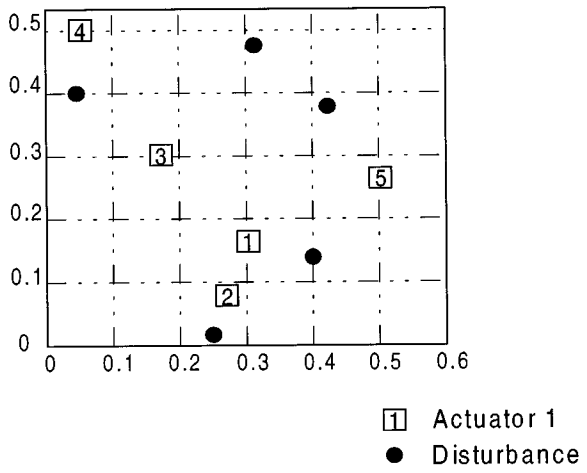


FIG. 3. Schematic diagram of plate with disturbance and actuator/sensor placement.

is a stationary point on the loss function, i.e., $dJ/d\mathbf{G}=0$. Therefore, $\mathbf{G}_{\text{opt}}$ satisfies the necessary conditions for optimality.

The procedure for calculating $\mathbf{G}_{\text{opt}}$ involves first determining an initial gain matrix, $\mathbf{G}_0$, which stabilizes the system. Equation (16) is then solved for $\mathbf{L}$, and the nonlinear equations,

$$0=\mathbf{K}\mathbf{A}_c+\mathbf{A}_c^T\mathbf{K}+\mathbf{C}_s^T\mathbf{G}^T R\mathbf{G}\mathbf{C}_s+\tilde{\mathbf{Q}}, \tag{18}$$

$$\mathbf{G}=\mathbf{G}_0+\theta(R^{-1}\mathbf{B}^T\mathbf{K}\mathbf{L}\mathbf{C}_s^T(\mathbf{C}_s\mathbf{L}\mathbf{C}_s^T)^{-1}-\mathbf{G}_0), \tag{19}$$

must be solved for $\mathbf{G}$ and $\mathbf{K}$. The scalar parameter θ is constrained to $0<\theta<1$, and can be adjusted downward to ensure that $J(\mathbf{G})<J(\mathbf{G}_0)$. The numerical algorithm proceeds by iterating on Eqs. (16), (18), and (19), and updating $\mathbf{G}$ until $|\mathbf{G}_{i+1}-\mathbf{G}_i|<\epsilon$. In practice, $\mathbf{G}_0=0$ was used as the initial stabilizing gain matrix and $\mathbf{P}$ was chosen to be $\mathbf{B}_d^T\mathbf{B}_d$, where $\mathbf{B}_d$ described a disturbance input influence matrix. Convergence of the algorithm on a 200th-order model typically occurred within 25 iterations.

III. PARAMETERS OF THE MODEL

Future experimental implementations of the control system are planned, and hence the model parameters are based upon that of a test structure used in the Adaptive Systems and Structures Laboratory at Duke University. A schematic diagram of the structure is illustrated in Fig. 3. Pinned boundary conditions are assumed and dimensions are such that $L_x=0.6$ m, $L_y=0.525$ m, $h_s=0.0016$ m. Material properties are consistent with that of steel with a Young's modulus: $E=19.5\times10^{10}$ Pa, material density: $\rho_s=7700$ kg/m^3, and proportional damping ratio: $\eta_s=0.05$. The speed of sound in the fluid (air) and fluid density were chosen at typical atmospheric conditions, respectively: $c=340$ m/s and $\rho_a=1.21$ kg/m^3.

An array of five point forces was used to generate exogenous inputs to the structure, and the position of each point force disturbance is depicted by the circles on the schematic diagram of Fig. 3. The disturbance inputs to each of the point forces were excited independently, and the positions of the point forces were randomly chosen to provide excitation of all plate modes over the desired bandwidth. The relative positions of the colocated point force actuator and point velocity sensor pairs are illustrated in Fig. 3 by square marks. Each transducer pair is numbered sequentially in Fig. 3 to convey the order in which they were added as the number of transducers used in the control system array were increased (i.e., 3I3O corresponds to positions 1, 2, and 3).

IV. RESULTS

Results presented in this section detail the performance of static compensation and compare this with state feedback control for a SISO, 3I3O, and 5I5O control system with colocated transducers. The dimension of the colocated array, the control effort penalty, and the effect of model order are studied parametrically to give insight into the physics and provide practical strategies. For each test case presented, the rms sound power radiated by the structure and the rms structural response to a random excitation are plotted as a function of frequency.

A. Number of colocated transducer pairs

For the results presented in this section, the radiation filters were used to generate the performance penalty while the control effort penalty was set to $R=10^{-5}$. This level of control effort penalty was chosen since it provided reasonable control gains and associated control forces.

In the limiting case where there are as many linearly independent sensors as there are states in the model the optimal static feedback solution becomes the state feedback solution and the two approaches are equivalent. For the cases considered here the augmented model has 216 states. Clearly this many sensors is not reasonable, however the underlying physics of the problem are not this complex and it is interesting to examine the performance of the static controller as the transducer array is increased in dimension.

The open-loop performance is compared with both the static and state feedback controllers in Fig. 4, for the case of 1, 3, and 5 transducers. The plots show the rms far-field acoustic power at each frequency from a random disturbance. Greater levels of attenuation are achieved for the high-frequency radiation modes with the state feedback controller, especially for the single actuator case. However, the overall performance of the static compensator is of a comparable level, which is encouraging from a practical perspective.

In Fig. 5 the rms acoustic power, rms structural power, and the normalized cost are plotted for the static controllers. The plots are parameterized by the number of transducers, (i.e., 0, 1, 3, 5 with 0 transducers being the open-loop re--sponse). As can be seen from the bar chart of the normalized cost, a significant gain is made upon increasing the number of transducers sequentially from 1 to 5. These observations are also reflected in the plots of the rms acoustic power and the rms structural power in each case. Notice the significant

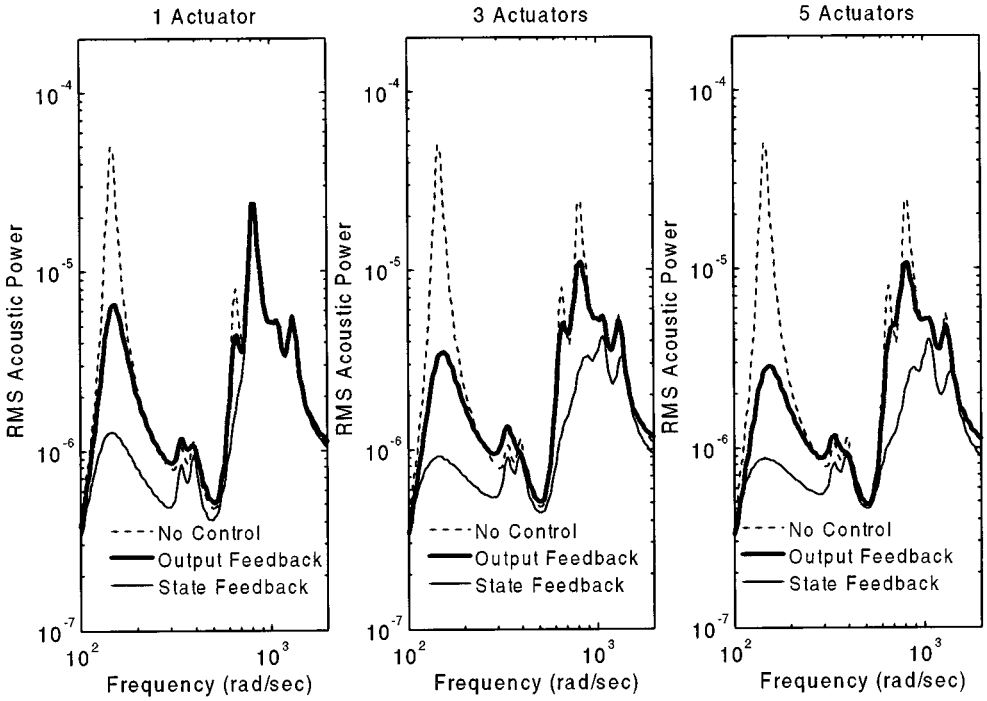


FIG. 4. Comparison of closed-loop performance: static compensation and state feedback compensation for SISO, 3I3O, and 5I5O controllers.

difference in the rms acoustic power and rms structural power at resonant frequencies corresponding to the (2,1) and (1,2) modes of the structure (between 300 and 500 rad/s). Over this bandwidth, these modes are inefficient radiators and as such do not contribute significantly to the rms acoustic power. Notice also that the static compensator did not increase the response of high-order structural modes, even though these modes do not drive the performance output. This result is due to the dissipative nature of the control law. The benefits of dissipative controllers will be addressed in more detail in Sec. II C 2.

B. Varying the control effort penalty

Results presented in this section assume three transducers are available and vary the control effort penalty an order of magnitude above and below the level considered in Sec. IV A. Increasing R beyond this range severely limits the

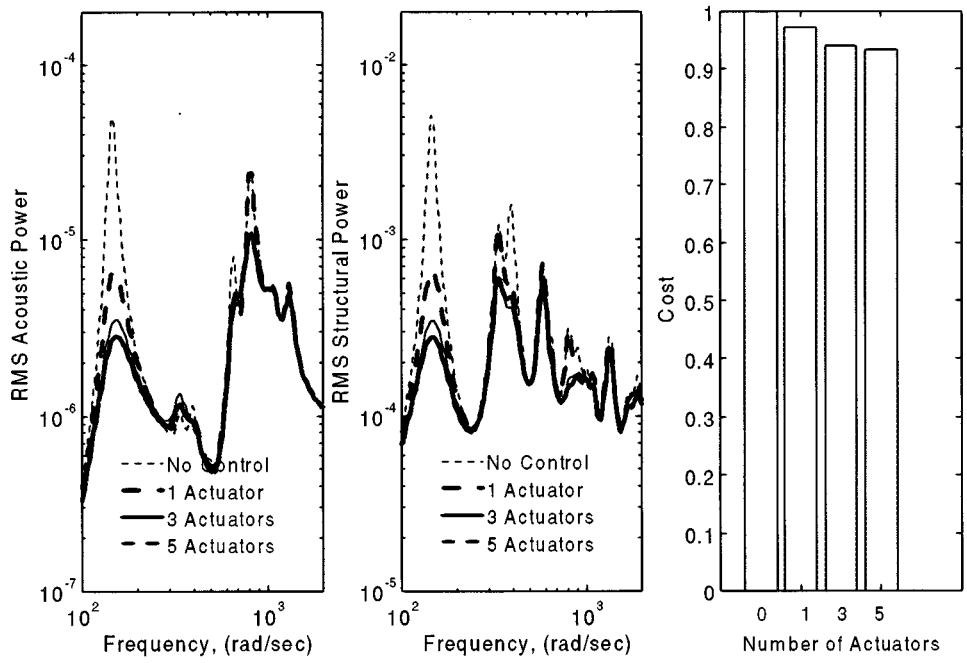


FIG. 5. Comparison of closed-loop performance with static compensation for SISO, 3I3O, and 5I5O controllers.

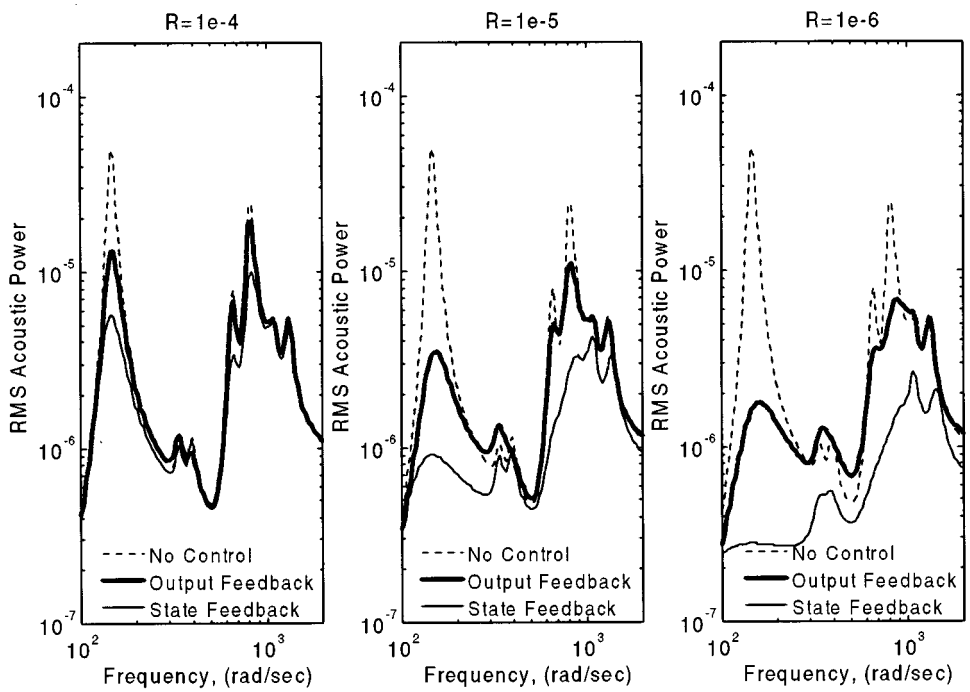


FIG. 6. Comparison of closed-loop performance: 3I3O static compensation and state feedback compensation for control effort penalties of 10^{-4}, 10^{-5}, and 10^{-6}.

level of control effort which can be applied and performance is essentially open loop. Lower levels of R lead to unreasonably large control forces.

As indicated in Fig. 6, for $R=10^{-4}$, very little difference is observed between the performance of the state feedback controller and the static compensator. For a control effort penalty of $R=10^{-5}$, the performance of the state feedback controller begins to differentiate itself from that of the static compensator. Upon reducing the control effort penalty to $R=10^{-6}$, the state feedback controller demonstrates much better performance over the bandwidth, and in the limit of no penalty on control effort, the state feedback controller can force a flat response over the entire bandwidth. Although arbitrary pole placement is a known advantage of state feedback, from a practical viewpoint it usually requires an unacceptable amount of control authority and high model fidelity.

In Fig. 7, plots of the rms acoustic power, rms structural

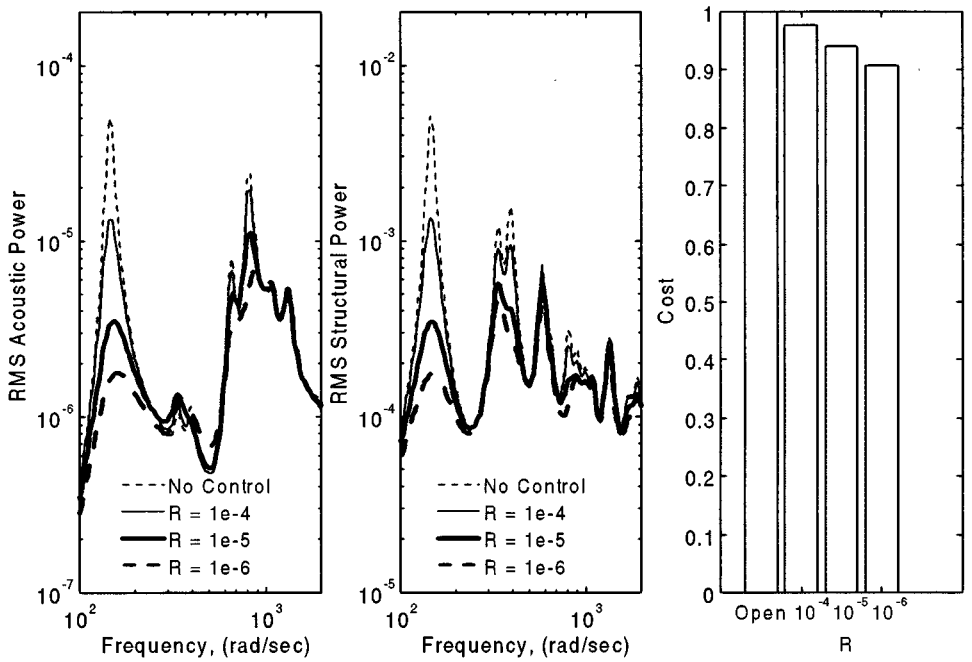


FIG. 7. Comparison of closed-loop performance with static compensation for a 3I3O controller and control effort penalties of 10^{-4}, 10^{-5}, and 10^{-6}.

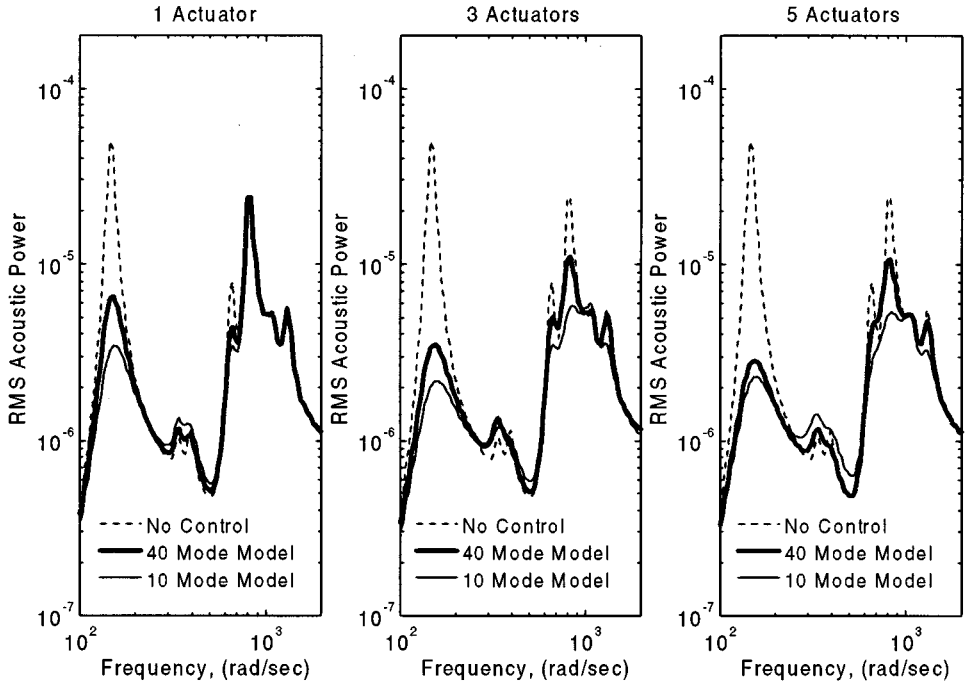


FIG. 8. Comparison of closed-loop performance with static compensation for SISO, 3I3O, and 5I5O controllers using 10 structural modes in one case and 40 structural modes in the other while implementing radiation filters for 10 modes in both cases and $R=10^{-5}$.

power, and normalized cost are presented for the static controller. From the normalized cost, one observes that the performance of the closed-loop system increases with decreasing levels of control effort penalty, as expected. The improvement in performance is also readily observed from the plot of the rms acoustic power, where it can be seen that the static controller is getting most of the reduction through attenuation of the first radiation mode. Similar trends are observed in the rms structural power, with attenuation being concentrated on the first, fourth and fifth modes, and less on the second and third modes, which are poor radiators. Again structural modes above the tenth, although not considered in the cost function, are also reduced slightly due to the fact that the static compensator resulted in a dissipative controller.

C. Effects of model order on control system synthesis

Practical limitations with respect to computational resources impose constraints on the model order which can be used in the design of feedback control systems. However, understanding the physics of the dynamic system guide the designer in choosing an appropriate number of states to include. For the problem under consideration, finite-order models need to be constructed for both the structural model and the radiation filters. For the structural acoustic control system, one must select enough states for the structural modes to accurately predict *closed-loop* performance. The radiation filters were constructed through curve fits of numerical data, and required an accuracy commensurate with the structural model fidelity over the bandwidth of interest. The effects of model order on system performance are discussed in this section.

1. Effects of the structural model order

As outlined by Clark (1995), the cumulative compliance of out-of-bandwidth modes plays a critical role in the prediction of closed-loop performance for gain feedback SISO control systems with colocated transducers. Out-of-bandwidth modes refers to modes of the structural model which are not considered important for the acoustic problem, and therefore do not drive the radiation filters. The velocity observations for the SISO case extend to the work presented herein for the static controller. Performance of the state feedback controller, because it assumes full knowledge, is not effected by the inclusion of out-of-bandwidth modes.

Consider the predicted closed-loop performance of a SISO, 3I3O, and 5I5O control system presented in Fig. 8. The control effort penalty was set to $R=10^{-5}$ and the performance penalty was as outlined for the prior test cases, with radiation filters for the first ten structural modes. The number of structural modes included in the dynamic model, however, was set to 10 and 40 for the two cases considered. For each case presented in Fig. 8, the open-loop rms acoustic power is compared against the closed-loop rms acoustic power. As indicated, whether 1, 3, or 5 colocated transducers were employed, the predicted performance for the ten-mode model was better than the performance predicted by the 40-mode model. This improvement in performance, however, is strictly an artifact of the simulation and would not be realized in practice because of the plant's true high-order dy-

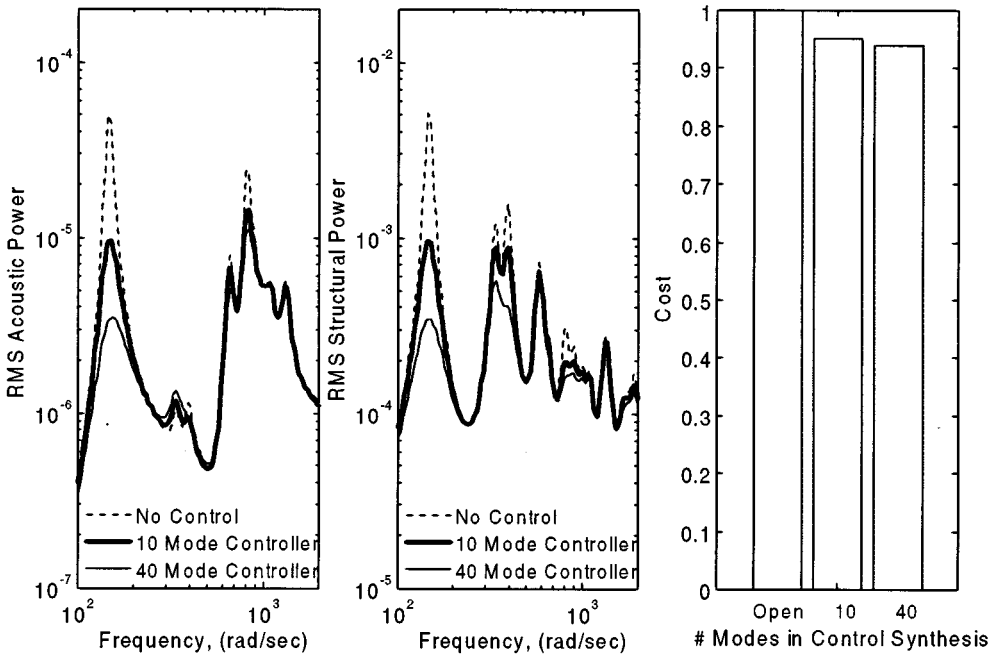


FIG. 9. Comparison of closed-loop performance for static compensation designed for a 10 mode model and a 40 mode model with both controllers implemented on a 40 mode model using 3 control actuators, radiation filters for 10 modes and $R=10^{-5}$.

namics. Increasing the structural model order beyond 40 modes yielded an insignificant variation in the predicted performance. The results demonstrate the importance of including the dynamics of structural modes well out of the desired bandwidth of analysis.

2. Dissipative controllers

To see the effects of unmodeled dynamics, the 3I3O static control law designed above for the system with 10 structural modes was applied to a 40-mode structural model. The gain matrix was positive definite, therefore this colocated velocity feedback system was dissipative (Balas, 1978). The dissipative nature of the controller provides excellent robustness and guarantees the system will remain stable even with unmodeled modes. Figure 9 shows the 10 mode controller implemented on a 40-mode model and compares this with the controller designed for the 40-mode model. In the plot of cost, open is used to represent the open-loop cost while 10 and 40 are used to represent the number of structural modes used in the design of the static compensator. All results presented in Fig. 9 use a 40 mode structural model in the computation of the cost. Although performance is sacrificed when a control law is designed with an inadequate system model, the dissipative controller remains stable, and even tends to attenuate the high-order structural modes. Since modally truncated models are typically used in the design process the dissipative control laws have significant practical advantages.

V. CONCLUSIONS

A formal procedure for designing static compensators for broadband structural acoustic control was developed. Results from this study demonstrate that direct output feedback control with static compensation can be used to effectively attenuate the sound radiation from a vibrating structure below coincidence, offering a convenient and physically realizable method of implementing a broadband, feedback, structural acoustic controller. As indicated in the section devoted to the results, for reasonable levels of control effort penalty ($10^{-4}<R<10^{-5}$ for the example presented), the performance of the static compensator was comparable to that of full state feedback over the low-frequency bandwidth of interest. In general, the performance of the static compensator can be increased by increasing the number of transducers, and in the limit of equal transducers and equal states, the static compensator can be used to achieve equivalent performance to the state feedback controller although this order of system would be unacceptable. For the low authority controller, performance gains obtained with static compensation neared that achieved with state feedback with only a limited number of transducers, even on a high-order system model. As was expected, relaxing the penalty on the control effort yields greater performance for both static compensation and state feedback compensation with more significant performance gains in the case of state feedback control.

In terms of model order, it was demonstrated that one must include a significant number of out-of-bandwidth structural modes to accurately capture the compliance of the colocated transducers for predicting closed-loop performance of the static compensator. In general, if an insufficient number of structural modes are included in the dynamic model,the closed-loop performance of the output feedback controller will be overpredicted. For state feedback control, the number of structural modes is relatively unimportant since full information is available in the control system design. Furthermore, the number of structural modes passed through the radiation filters imposed a less significant error in the pre-

dicted performance of the closed-loop system with static compensation than with state feedback. The reason for this observation is that all of the static compensators designed and implemented in this study resulted in positive definite feedback gains and thus were dissipative control systems. The dissipative controller is guaranteed stable and has relatively little effect on out-of-bandwidth modes. This can be a significant advantage in structural system systems which have high modal density at frequencies where reverberant models fail to capture the dynamics.

ACKNOWLEDGMENT

This work was supported in part by NASA Langley Research Center under Grant No. NAG-1-1570.

Anderson, B., and Moore, J. (**1990**). *Optimal Control: Linear Quadratic Methods* (Prentice-Hall, Englewood Cliffs, NJ).

Balas, M. J. (**1978**). ''Direct VElocity Feedback Control of Large Space Structures,'' J. Guid. Control **2**, 252–253.

Baumann, W. T., Saunders, W. R., and Robertshaw, H. H. (**1991**). ''Active suppression of acoustic radiation from impulsively excited structures,'' J. Acoust. Soc. Am. **90**, 3202–3208.

Baumann, W. T., Ho, Fu-Sheng, and Robertshaw, H. H. (**1992**). ''Active structural acoustic control of broadband disturbances,'' J. Acoust. Soc. Am. **92**, 1998–2005.

Borgiotti, G. (**1990**). ''The power radiated by a vibrating body in an acoustic fluid and its determination from boundary measurements,'' J. Acoust. Soc. Am. **88**, 199–209.

Burke, S. E., and Hubbard, J. E., Jr. (**1990**). ''Spatial filtering concepts in distributed parameter control,'' J. Dyn. Syst. Measurement and Control **112**, 565–573.

Clark, R. L. (**1995**). ''Accounting for out-of-bandwidth modes in the assumed modes approach: Implications on colocated output feedback control,'' J. Dyn. Syst. Measurement Control (accepted for publication).

Cunefare, K. A. (**1991**). ''The minimum multimodal radiation efficiency of baffled finite beams,'' J. Acoust. Soc. Am. **90**, 2521–2529.

Francis, B. (**1986**). *A Course in H_∞ Control Theory* (Springer-Verlag, New York).

Fuller, C. R. (**1988**). ''Analysis of active control of sound radiation from elastic plates by force inputs,'' Proceedings of Inter-Noise 88, Avignon, **2**, 1061–1064.

Gupta, N. K. (**1980**). ''Frequency shaped cost functionals: extension of linear-quadratic Gaussian design methods,'' J. Guid. Control **3**, 529–535.

Kalman, R. E., Ho, Y., and Narendra, K. S. (**1963**). ''Controllability of linear dynamical systems,'' in *Contributions to Differential Equations*, Vol. 1 (Wiley, New York).

Levine, W. S., and Athans, M. (**1970**). ''On the determination of the optimal constant output feedback gains for linear multivariable systems,'' IEEE Trans. Autom. Control. **AC-15**, 44–48.

Lomas, N. S., and Hayek, S. I. (**1977**). ''Vibration and acoustic radiation of elastically supported rectangular plates,'' J. Sound Vib. **52**, 1–25.

Makila, P. M., and Toivonen, H. T. (**1987**). ''Computational methods for parametric LQ problems—A survey,'' IEEE Trans. Autom. Control. **AC-32**, 658–671.

Meirovitch, L. (**1967**). *Analytical Methods in Vibrations* (Macmillan, New York).

Photiadis, D. (**1990**). ''The relationship of singular value decomposition to wave vector filtering in sound radiation problems,'' J. Acoust. Soc. Am. **88**, 1152–1159.

Thomas, D. R., and Nelson, P. A. (**1995**). ''Feedback control of sound radiation from a plate excited by a turbulent boundary layer,'' J. Acoust. Soc. Am. **98**, 2651–2662.

Toivonen, H. T. (**1985**). ''A globally convergent algorithm for the optimal constant output feedback problem,'' Int. J. Control **41**, 1589–1599.

Array technology for acoustic wave field analysis in enclosures

A. J. Berkhout, D. de Vries, and J. J. Sonke
Laboratory of Seismics and Acoustics, Department of Applied Physics, Delft University of Technology, 2600 GA Delft, The Netherlands

(Received 26 July 1996; revised 10 July 1997; accepted 30 July 1997)

A method is proposed to calculate and measure impulse responses in an enclosure along closely spaced receiver arrays. Hence, instead of using a sparse distribution of receiver positions with single microphones, as is common practice now, *arrays* of microphones are applied to register the complex sound fields within enclosures. This way, there is a strong spatial correlation between adjacent responses, enabling one to analyze individual reflected wavefronts. It is shown that visualization of the recorded data in a two-dimensional domain, defined by detector position and travel time, gives a significantly improved insight in the structure of complex sound fields. This insight is further increased by applying the linear Radon transform (plane wave decomposition), yielding a representation of the data in the so-called ray parameter versus intercept time domain. © *1997 Acoustical Society of America.* [S0001-4966(97)04611-0]

PACS numbers: 43.55.Br, 43.60.Gk [JDQ]

INTRODUCTION

Since the early 60s, sound fields in enclosures have been characterized by impulse responses, computed and measured for a limited number of representative source and receiver positions. The choice of these positions is rather arbitrary. Physical intuition, supported by perceptual experiments, has established the general opinion that the fine structure of these responses is highly redundant from a perceptual point of view; see Fig. 1. As a consequence, the individual microphone signals are analyzed in a *statistical* way only, yielding a number of widely used energy-based characterizations.[1,2] However, when microphone arrays are used then statistical analysis can be replaced by *wave field analysis* and the amount of insight and information that can be extracted is significantly increased. Note that in an earlier paper by Berkhout *et al.*[3] the integration of array measurements and wave field analysis has been proposed for outdoor sound fields.

In this paper, linear arrays of closely spaced microphones are considered that record the response of a single point source in an enclosed space. In the plane through source and array (the acquisition plane) we define a two-dimensional Cartesian coordinate system (ξ,η) having its origin at the source position; see Fig. 2.

The ξ axis is parallel to the microphone array, the η axis intersects the array perpendicularly at $(\xi=0,\ \eta=\eta_a)$. Hence, the position of a microphone is given by (ξ_i,η_a); ξ_i is called the *offset* of the i-th microphone. As will be derived later, the free-field response of the source at the array in the offset versus travel time domain has a hyperbolic shape. The apex of the hyperbola is found at $(\xi=0,\ t=\eta_a/c)$. An example of the offset-travel time response measured in an enclosure is given in Fig. 3. Reflections appear as hyperbolic, parabolic, and linear wavefronts and can be easily identified, thus giving insight into the spatial properties of the wave field.

Valuable information on the spatial properties of the wave field can be extracted when the data in the recording domain, i.e., the (ξ,t) domain, are Radon transformed to the so-called ''ray parameter versus intercept time (p,τ) domain.'' The Radon transform will be introduced in Sec. I. The free-field array response of a point source in both the (ξ,t) and the (p,τ) domain will be discussed in Sec. II. In Sec. III, simulated array responses of a source in an enclosure will be analyzed in both domains. In Sec. IV, this analysis is applied to measured array responses. In Sec. V it is shown that array measurements in the (p,τ) domain are very suitable to select wave field components reflected at specific boundaries. By applying an inverse Radon transform, these components can be studied in the (ξ,t) domain. Finally, in Sec. VI conclusions and topics of current research are formulated.

Considering an enclosed space as a time-invariant system, at least during one measurement session, then array measurements can be obtained with a single microphone by moving the microphone along the array line, recording at *each* desired position one impulse response. Nowadays, software driven measurement systems are available that transport transducers with high precision along a rail and fully automatically carry out predesigned measurement programs without any user interaction. We expect that array technology will significantly contribute to the improvement of measurement and analysis methods in applied room acoustics.

I. PLANE WAVE RESPONSE AT THE MICROPHONE ARRAY

We consider a plane wave propagating in the (ξ,η) plane [Fig. 4(a)] that is incident on the array under an angle α [Fig. 4(b)].

In the space–time domain the pressure of this wave is written as

$$\mathrm{p}(\xi,\eta,t)=s\left(t-\frac{\xi}{c_\xi}-\frac{\eta}{c_\eta}\right)=s(t-p\xi-q\eta), \tag{1}$$

where $s(t)$ denotes the source signature. Figure 4(b), showing two snapshots of the wavefront at times t and $t+dt$,

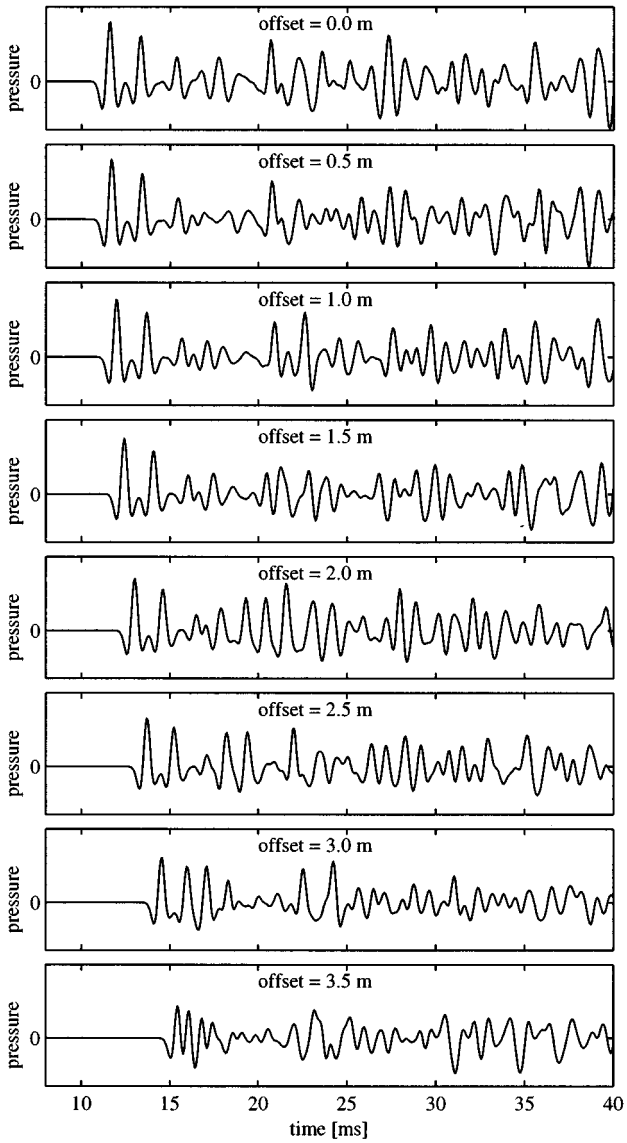


FIG. 1. Impulse responses of one and the same source, measured at eight microphone positions 0.5-m apart. See also Fig. 3. Note the significant differences.

illustrates that c_ξ and c_η represent the phase velocities in the ξ and η direction, respectively,

$$c_\xi = \frac{d\xi}{dt} = \frac{c}{\sin\alpha} \tag{2a}$$

and

$$c_\eta = \frac{d\eta}{dt} = \frac{c}{\cos\alpha}. \tag{2b}$$

The *ray parameters* p and q are defined as the inverse of the phase velocities and hence to be interpreted as the horizontal and vertical phase *slowness* of the plane wave, respectively,

$$p = \frac{1}{c_\xi} = \frac{dt}{d\xi} = \frac{\sin\alpha}{c} \tag{3a}$$

and

$$q = \frac{1}{c_\eta} = \frac{dt}{d\eta} = \frac{\cos\alpha}{c}, \tag{3b}$$

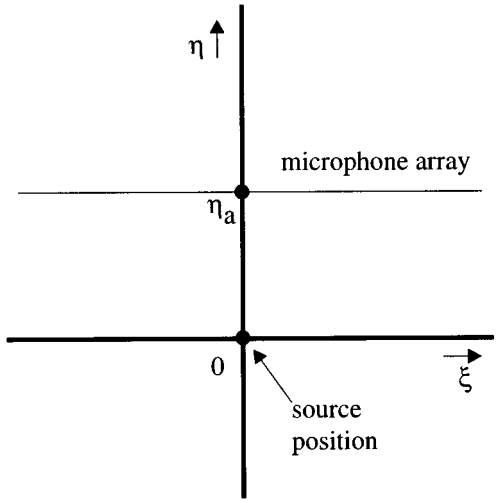


FIG. 2. Coordinate system (ξ,η) in the acquisition plane, with the source at the origin and the microphone array at $\eta=\eta_a$.

such that

$$p^2+q^2=c^{-2}. \tag{4}$$

Equation (3) shows that, for a given propagation velocity, the ray parameters p and q unambiguously correspond with an angle of incidence α.

Now we consider the pressure of a specific plane wave with $p=p_i$, $q=q_i$ at the microphone array, i.e., for $\eta=\eta_a$. According to Eq. (1) we may write

$$\mathrm{p}(\xi,\eta_a,t)=s(t-p_i\xi-q_i\eta_a), \tag{5a}$$

or, substituting

$$\tau_i=q_i\eta_a, \tag{5b}$$

$$\mathrm{p}(\xi,\eta_a,t)=s(t-p_i\xi-\tau_i), \tag{5c}$$

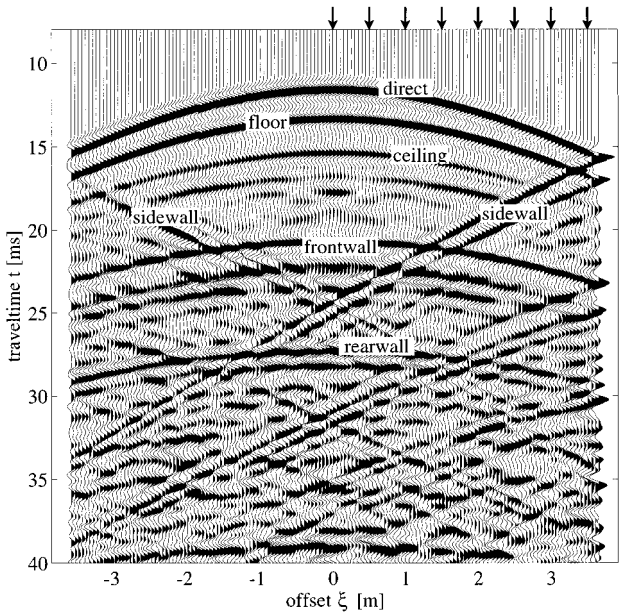


FIG. 3. Offset versus travel time (ξ,t) representation of the soundfield in an enclosure that is recorded with a closely spaced microphone array ($\Delta\xi=0.05$ m). The impulse responses shown in Fig. 2 are indicated with arrows.

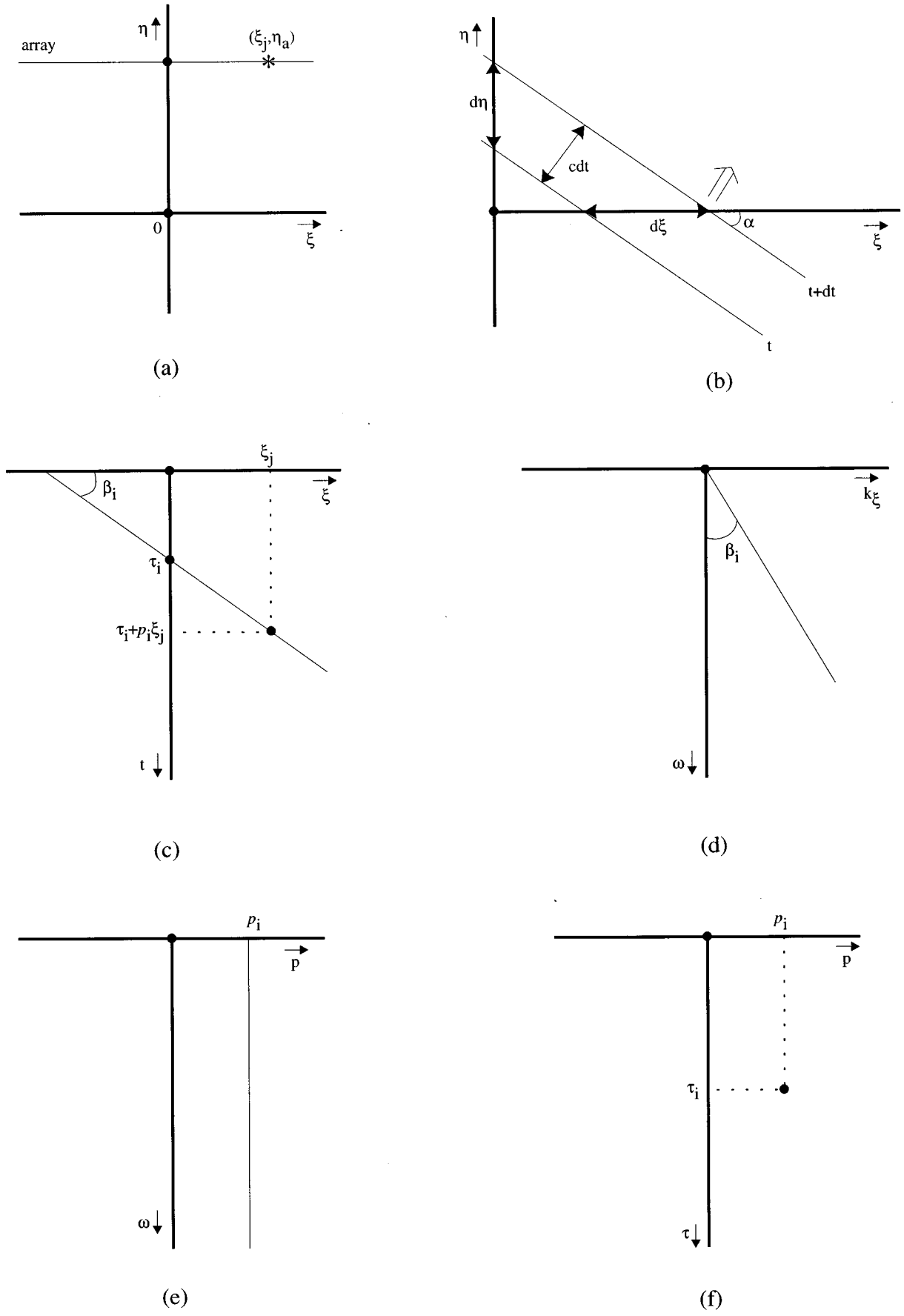


FIG. 4. Plane wave response at the array. (a) Configuration and coordinates; (b) representation of the response in the space domain (two wavefronts); (c) id. in the offset versus travel time domain; (d) id. in the wave number versus frequency domain; (e) id. in the ray parameter versus frequency domain; (f) id. in the ray parameter versus intercept time domain.

where τ_i denotes the time it takes for the wave to travel from the origin of the coordinate system to the zero-offset microphone of the array ($\xi=0,\ \eta=\eta_a$). This time is called the *intercept time* and, according to Eq. (3b), decreases with increasing angle of incidence α. If we take $t=0$ at the origin, then the arrival time of the wavefront at the zero-offset microphone equals τ_i. For the microphone with offset ξ_j the arrival time equals

$$t=\tau_i+p_i\xi_j. \tag{6}$$

A graphic representation is given in Fig. 4(c): In the offset-travel time (ξ,t) domain the pressure field of a plane wave at the array appears as a straight line with a slope given by $\tan\beta_i=p_i$. For full lateral incidence, i.e., along the ξ axis ($\alpha=\pi/2$), p reaches its maximum value at $c^{-1}\approx 3$ ms/m for air at room temperature, corresponding with a maximum value of β of about 3×10^{-3} rad.

Equation (5c) can be Fourier transformed to the offset-frequency domain, yielding

$$\begin{aligned} P(\xi,\eta_a,\omega)&=\int \mathrm{p}(\xi,\eta_a,t)e^{-j\omega t}\,dt\\ &=\int s(t-\tau_i-p_i\xi)e^{-j\omega t}\,dt\\ &=S(\omega)e^{-j\omega(\tau_i+p_i\xi)}, \end{aligned} \tag{7}$$

where $S(\omega)$ denotes the frequency spectrum of the source. When also a spatial Fourier transform is applied, the pressure field at the array is represented in the wave number frequency domain. Omitting the argument $\eta=\eta_a$ for notation simplicity, we write

$$\begin{aligned} \widetilde{P}(k_\xi,\omega)&=\int P(\xi,\omega)e^{jk_\xi\xi}\,d\xi\\ &=\int S(\omega)e^{-j\omega(\tau_i+p_i\xi)}e^{jk_\xi\xi}\,d\xi\\ &=S(\omega)e^{-j\omega\tau_i}\int e^{j(k_\xi-\omega p_i)\xi}\,d\xi\\ &=2\pi S(\omega)e^{-j\omega\tau_i}\delta(k_\xi-\omega p_i), \end{aligned} \tag{8}$$

where δ represents the delta pulse. Hence, the pressure of a plane wave in the double Fourier domain is mapped on the line $k_\xi=\omega p_i$; see Fig. 4(d). Instead of transforming the offset coordinate to the wave number coordinate ($\xi\rightarrow k_\xi$), it can also be transformed to the horizontal ray parameter coordinate ($\xi\rightarrow p$):

$$\begin{aligned} \widetilde{P}(p,\omega)&=\int P(\xi,\omega)e^{j\omega p\xi}\,d\xi\\ &=\int S(\omega)e^{-j\omega(\tau_i+p_i\xi)}e^{j\omega p\xi}\,d\xi\\ &=S(\omega)e^{-j\omega\tau_i}\int e^{-j\omega(p_i-p)\xi}\,d\xi\\ &=2\pi\frac{S(\omega)}{|\omega|}e^{-j\omega\tau_i}\delta(p-p_i). \end{aligned} \tag{9}$$

As expected, all frequency components of a traveling plane have the same ray parameter and can thus be found on a vertical line in the ray parameter domain; see Fig. 4(e). Finally, the pressure field at the array can be transformed to the ray parameter versus intercept time domain by applying the inverse Fourier transform ($\omega\rightarrow t$) to Eq. (9):

$$\begin{aligned} \widetilde{\mathrm{p}}(p,\tau)&=\frac{1}{2\pi}\int \widetilde{P}(p,\omega)e^{j\omega\tau}\,d\omega\\ &=\left[\int \frac{S(\omega)}{|\omega|}e^{j\omega(\tau-\tau_i)}d\omega\right]\delta(p-p_i)\\ &=s'(\tau-\tau_i)\,\delta(p-p_i), \end{aligned} \tag{10}$$

where $s'(\tau)$ and $S(\omega)/|\omega|$ are a Fourier pair. Equation (10) shows that one plane wave is transformed into a single pulse in the (p,τ) plane, the position of this pulse being given by the ray parameter p_i and the intercept time τ_i; see Fig. 4(f). This means that a distribution of plane waves in the offset versus travel time domain is transformed into a distribution of single pulses in the ray parameter versus intercept time domain.

Due to the use of finite microphone arrays, we always deal in practice with truncated plane waves, meaning that each delta function is broadened along the p coordinate, the width being inversely proportional to the width of the truncation window. Array length and temporal bandwidth thus determine the *resolution* in the (p,τ) domain.

From Eq. (9) it follows that the representation of the array measurements in the ray parameter versus intercept time domain can also be written as

$$\widetilde{\mathrm{p}}(p,\tau)=\int \mathrm{p}(\xi,\tau+p\xi)d\xi. \tag{11a}$$

Pressure signals in the offset-travel time domain are integrated (summed) along trajectories each defined by a p and a τ value. For example, plain superposition of all microphone signals (traces) in Fig. 3 means a summation of pressure along horizontal lines (ray parameter $p=0$) for each travel time t (which for $p=0$ equals intercept time τ), yielding the $p=0$ contribution in the ray parameter-intercept time domain.

The operation in Eq. (11a) is called the linear Radon transform,[4] often used in seismic signal processing and also known there as the ''τ-p transform'' or ''slant stack operation.'' In discrete form, the linear Radon transform reads

$$\widetilde{\mathrm{p}}(p_i,\tau_k)=\sum_j \mathrm{p}(\xi_j,\tau_k+p_i\xi_j)\Delta\xi. \tag{11b}$$

By applying this transform to data measured by a discrete array, the recorded wave field is thus decomposed into traveling plane waves, each plane wave being represented by a single response in the (p,τ) plane. Inversely, from a set of responses represented in the (p,τ) domain a wave field in the offset versus travel time domain can be composed by applying an inverse Radon transform, reading in the frequency domain

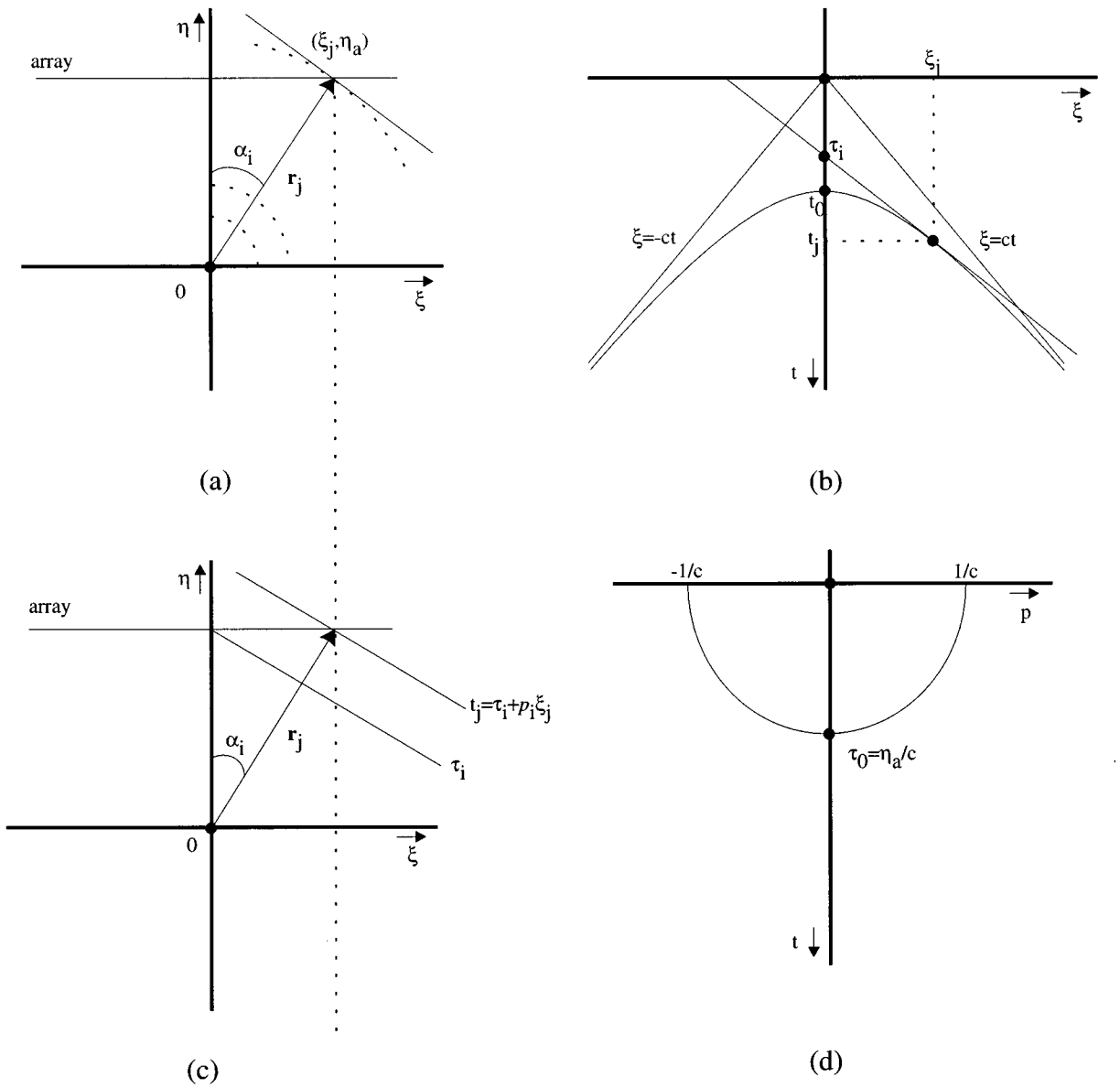


FIG. 5. Response of a point source at the array. (a) Configuration and coordinates; (b) representation of the response in the offset versus travel time domain (hyperbola); (c) plane wave component in the space domain (two wavefronts); (d) representation of the response in the ray parameter versus intercept time domain (ellipse).

$$P(\xi,\omega)=\frac{|\omega|}{2\pi}\int \tilde{P}(p,\omega)e^{-j\omega p\xi}\,dp. \tag{12}$$

II. POINT SOURCE RESPONSE AT THE ARRAY

We now consider the response at the array of a point source positioned at the origin of the coordinate system of Fig. 2. The distance between the source and a microphone with offset ξ_j is called r_j. From the configuration in Fig. 5(a) the expression for the involved travel distances can be easily derived:

$$r_j^2=\xi_j^2+\eta_a^2. \tag{13}$$

Substituting $t_0=\eta_a/c$ (zero-offset travel time) and $t_j=r_j/c$ (offset-travel time), the expression for the involved travel times follows directly from Eq. (13):

$$\frac{t^2}{t_0^2}-\frac{\xi^2}{\eta_a^2}=1, \tag{14}$$

which defines a *hyperbola* with asymptotes $\xi=\pm ct$; see Fig. 5(b). It means that the arrival times (travel times) of a point source response, measured along a linear array, are defined by a hyperbola.

Given a dipole source *on* the array at $(0,\eta_a)$, we may write for the pressure field along the microphone array

$$P(\xi,\omega)=S(\omega)\,\delta(\xi). \tag{15}$$

According to Eq. (9) its representation in the (p,τ) domain reads

$$\tilde{P}(p,\omega)=S(\omega)\int \delta(\xi)e^{j\omega p\xi}\,d\xi=S(\omega). \tag{16}$$

This means that the wave field of a dipole can be seen as a distribution of plane waves with equal spectral strength in each direction (for each p value). Figure 5(c) displays one of these plane waves incident on the array with angle α_i, due to a dipole source at the origin. According to the equations derived in the previous section we may write

$$\mathrm{p}(\xi,t)=s(t-p_i\xi-\tau_i) \tag{17a}$$

with

$$\tau_i=\frac{\cos\ \alpha_i}{c}\ \eta_a \tag{17b}$$

and

$$p_i=\frac{\sin\ \alpha_i}{c}. \tag{17c}$$

Combining the above equations, and omitting again the index i, we write for the relation between horizontal ray parameter and intercept time

$$\frac{\tau^2}{\tau_0^2}+\frac{p^2}{(1/c)^2}=1, \tag{18}$$

where $\tau_0=\eta_a/c=t_0$. Equation (18) represents an *ellipse* with half-axes $1/c$ and τ_0; see Fig. 5(d). It means that the arrival times of the point source plane wave components (intercept times) along a linear array are defined by an ellipse. We may conclude from Eqs. (14) and (18) that hyperbolic responses in the (ξ,t) domain are transformed to elliptic responses in the (p,τ) domain.

The raypath vector $\mathbf{r}_j$ between source position and receiver point with offset ξ_j, as indicated in Fig. 5(a), is always perpendicular to the tangent of the wavefront. This tangent is equivalent to the plane wave component defined by ray parameter p_i and intercept time τ_i, as shown in Fig. 5(c). This means that this plane wave appears as the tangent in point (ξ_j,t_j) at the hyperbola representing the data in the (ξ,t) domain; Fig. 5(b). This figure shows that we may write for the slope of this tangent

$$\left.\frac{dt}{d\xi}\right|_{\xi_j}=p_i=\frac{t_j-\tau_i}{\xi_j}, \tag{19}$$

which is identical with Eq. (6) describing a plane wave with ray parameter p_i in the (ξ,t) domain.

III. NUMERICALLY SIMULATED RESPONSES OF A POINT SOURCE IN AN ENCLOSURE

A. Representation in the offset versus travel time domain

The mirror image modeling method has been used to generate the simulations discussed in this section. Figure 6(a) shows the simulated wave field of a monopole at position A, as specified in the Appendix, recorded with a linear microphone array oriented in parallel with the front and back walls and extending over the full width of the hall. For this configuration, the zero-offset point is located at the array center. In this introductory example, the absorption coefficients of floor and ceiling have been chosen 1.0 such that effectively a two-dimensional room configuration results. Each individual time signal, measured at some offset ξ, represents the ''traditional'' impulse response at a certain receiver position, illustrating the local reflection sequence as a function of traveltime. Displaying the impulse responses for all neighbouring receiver positions (offsets) along the linear array reveals the spatial structure of the wave field. The first arrivals represent the hyperbolic direct wave front [marked ''1'' in Fig. 6(a)], as was theoretically discussed in Sec. II.

If the room is infinitely wide and the microphone array is infinitely long, the hyperbolic direct wavefront will extend unlimited in both lateral directions, approaching the asymptotes $\xi=\pm ct$. In a bounded room, however, the direct wavefront is reflected by the side walls. In the offset-travel time display, this appears as a ''folding'' of the hyperbola. In Fig. 6(a), the positions where the left-hand and the right-hand side of the hyperbola are folded, are indicated with an arrow. Detailed analysis shows that: (a) at the recording positions nearest to the side walls, the pressure amplitude increases due to the real and positive reflection coefficient; (b) each folded version has a lower amplitude than the preceding one, due to geometrical spreading during propagation and absorption during reflection; (c) after a few reflections, the folded versions appear as plane waves traveling in different directions.

More hyperbolic wavefronts can be distinguished in Fig. 6(a) (the first three are marked ''2,'' ''3,'' and ''4''), having their apices at later delay times and, hence, having a flatter shape. These hyperbolae represent spherical wavefronts first reflected by the front wall (''2''), the rear wall (''3''), first front then rear wall (''4''), etc. Note that, apart from travel time information, there is no discrimination between wavefronts arriving from front or rear as long as omnidirectional microphones are used. The reflections of these wavefronts at the side walls again appear as folded versions. It can be clearly observed that the density of these folded versions and the variety of their propagation directions increase with travel time. This phenomenon qualitatively supports the usual assumption that later reflections together form an isotropic distribution of plane waves.

Figure 6(b) shows the response for the same source and array positions, now with floor and ceiling having the absorption coefficients as given in the Appendix, making the configuration acoustically three dimensional. The reflection density is now much higher and increases much faster with travel time than in the 2-D case above. The plane waves representing the later reflections are seen to propagate in many directions, thus building up a reverberant field with a high degree of diffusivity. Note that the array cannot discriminate between different vertical angles of incidence. In the displays of Fig. 6, wavefronts having a vertical propagation component are seen as rotated to the horizontal plane over a cylinder with radius t. In order to analyze the vertical spatial properties of the wave field, vertical microphone arrays (or source arrays) should be applied. Figure 6(c) shows the offset-travel time response at the same array, but now with the source at position B as specified in the Appendix. For this configuration, zero offset does not coincide with the array centre. In terms of reflection density and degree of

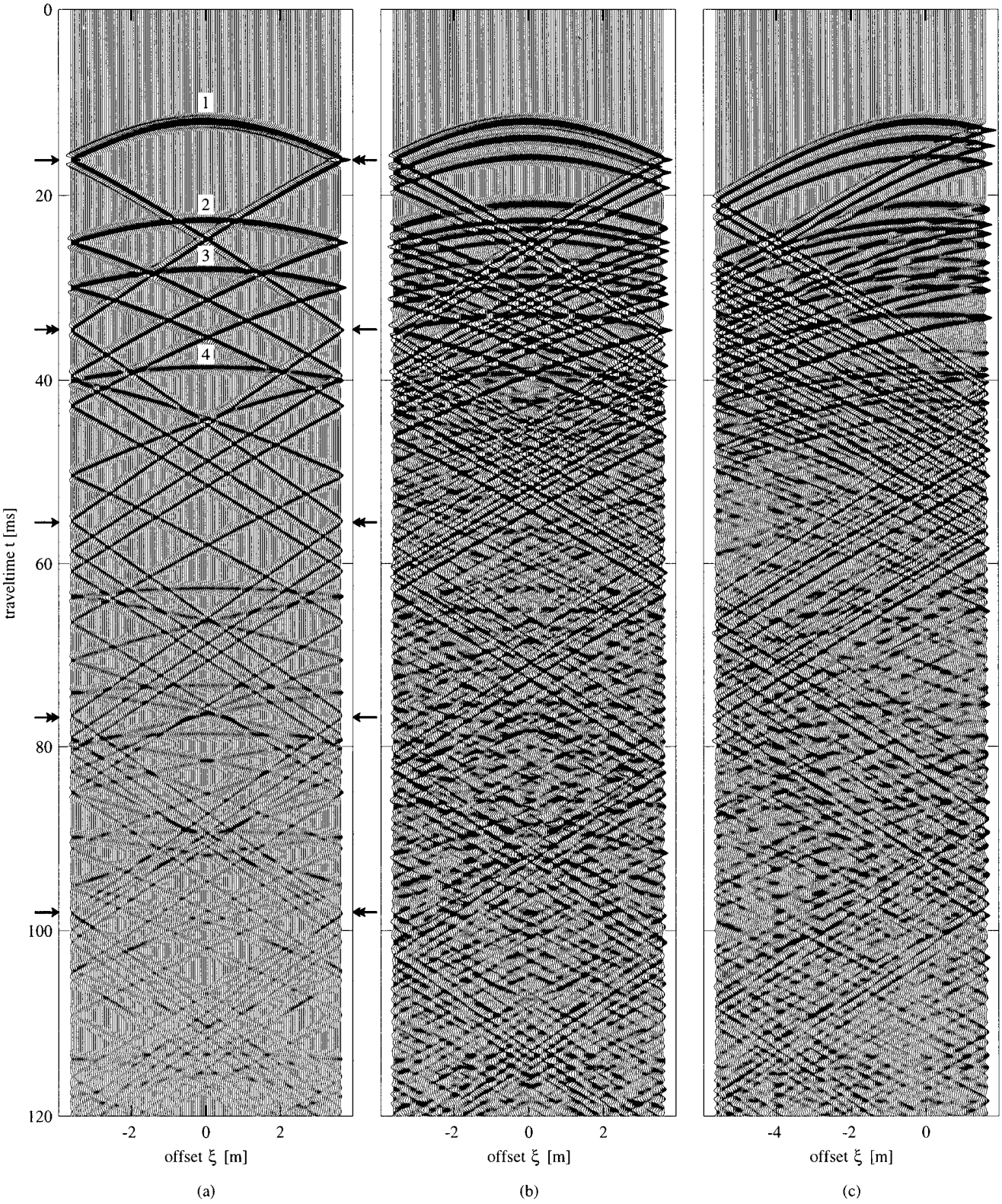


FIG. 6. Simulated point source responses in the offset versus travel time domain. (a) Zero offset at array center, two-dimensional situation; (b) id., three-dimensional situation; (c) zero offset right from array center, three-dimensional situation.

diffusivity, the wave field shows the same trends as that of Fig. 6(b).

It should be mentioned that, in the mirror-image simulation method considered here, only specular reflections are considered. Effects of diffraction due to the finite boundary dimensions and boundary irregularities are not taken into account. Such effects will always be present in practice, and will appear in the offset versus travel time domain as extra hyperbolic responses of secondary sources positioned at the boundaries.

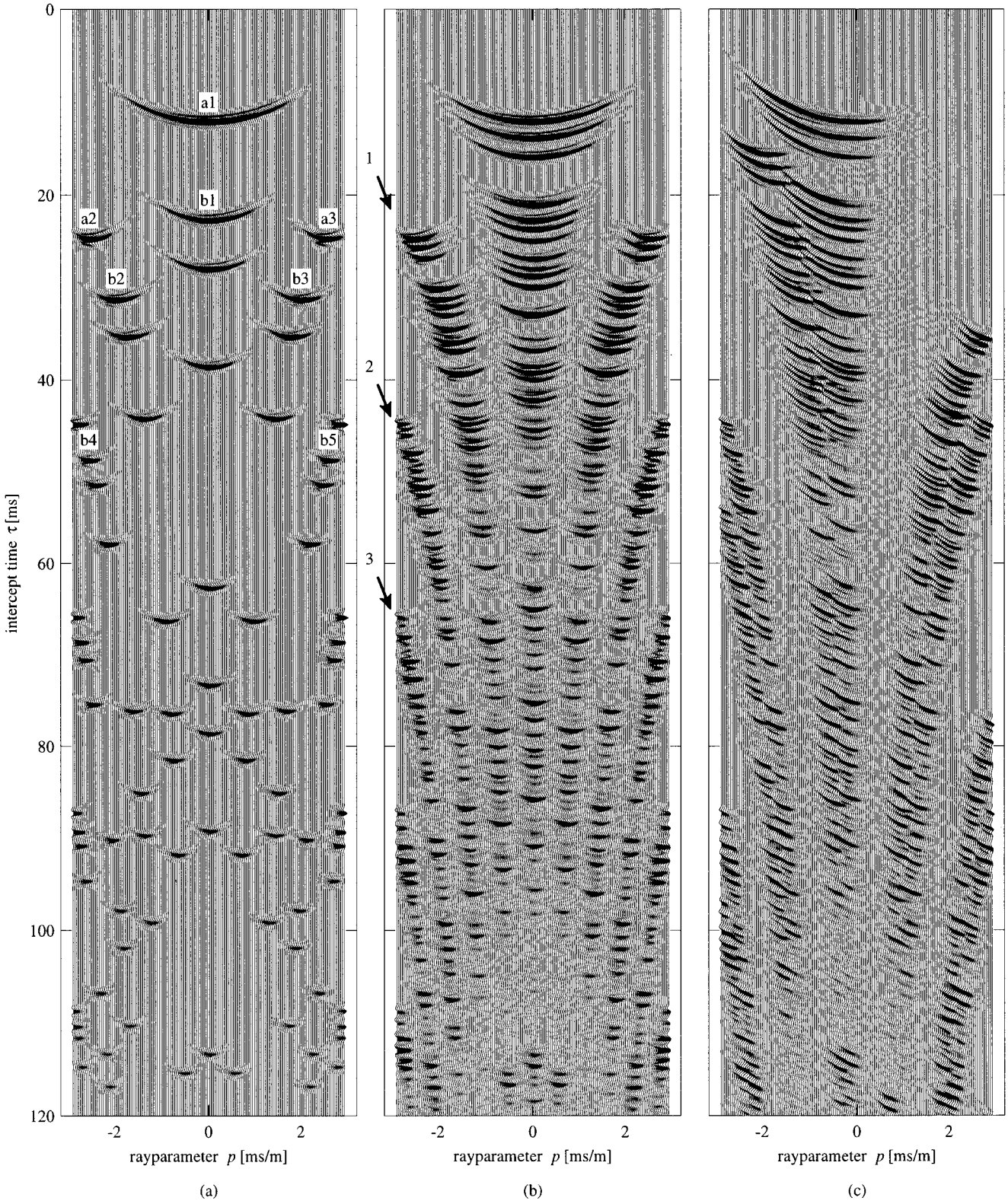


FIG. 7. Simulated point source responses in the ray parameter versus intercept time domain. (a) Zero offset at array center, two-dimensional situation; (b) id., three-dimensional situation; (c) zero offset right from array center, three-dimensional situation.

B. Representation in the ray parameter versus intercept time domain

In Fig. 7(a), the Radon transform of the 2-D offset-travel time response in Fig. 6(a) is given. The central trace gives the plane wave components with $p=0$, i.e., the tangents at the apices of the hyperbolae in the offset versus travel time domain, representing the first arrivals of the direct wavefront and the reflections at the front and rear walls. Note that for $p=0$ intercept time τ equals travel time t. Around the central trace, the Radon transform of the hyperbolae appears as

symmetric elliptic curves which, with increasing τ, decrease in width. This corresponds with the fact that, in the (ξ,t) domain, the hyperbolae become flatter with increasing delay time, thus better approximating plane waves which appear as bandlimited delta functions in the (p,τ) domain.

It is seen that the upper elliptic curve [marked 'a1' in Fig. 7(a)], being the Radon transform of the direct wavefront recorded by the microphone array, ends at a certain value of $|p|$, corresponding with the slope of hyperbola ''1'' of Fig. 6(a) at its outer points, i.e., near the side walls. The wavefronts reflected once at the side walls, appearing as folded hyperbola sections in Fig. 6(a), show up as the small elliptic curves marked ''a2'' (having negative p values and thus found at the *left-hand* side of the display, but representing the *right-hand* sidewall reflection) and ''a3'' (found at the *right-hand* display side, representing the *left-hand* sidewall reflection) in Fig. 7(a). Due to the specular character of the reflections, the outer p values of ''a1'' link up with the inner p values of ''a2'' and ''a3'': When the wavefront is reflected, the slope just changes sign. Since, in Fig. 6(a), the tangents at the outer points of the reflected wavefronts intersect the zero offset trace ($\xi=0$) at a different time than the tangents at the outer points of the direct wavefront, there is a discontinuity in the τ coordinates. In the same way, the ellipse ''b1,'' representing the first reflection at the front wall, is related to curves ''b2''–''b5'' which represent its first- and second-order reflections at the side walls.

Figure 7(b) shows the Radon transform of the offset-travel time data displayed in Fig. 6(b), related to the acoustically three-dimensional situation. It is now clearly seen that all sidewall reflections of the same order are seen to form data clusters in the (p,τ) domain, tending to a narrower area at smaller p values for increasing intercept times. Due to the symmetry of the configuration, the onset of the bands for increasing reflection order is periodic in τ. These effects can be easily understood by considering the simple two-dimensional image source model of Fig. 8. The column marked ''1'' (''2,'' ''3,''....) contains all mirror-image sources the response of which has reflected once at the right-hand sidewall (left and then right, right-left-right,....) and one or more times at the front and/or rear wall, corresponding with data clusters in band ''1'' (''2,'' ''3,''....) of Fig. 7(b). Note that in order to include reflections at floor and ceiling as well, a three-dimensional image source model is needed. Indeed, it is seen that the further the image source is removed from the array—i.e., the more often the wavefront has reflected against the front and back wall—the smaller the angle of incidence on the array, corresponding with decreasing p values for increasing τ values in the (p,τ) domain. The fact that, for increasing intercept times, the Radon transform of the wave field more and more breaks up into bandlimited delta functions at an increasing variety of p values clearly illustrates the increasing diffusivity of the reflected wave field with increasing time, already discovered in the (ξ,t) displays of Fig. 6. Figure 7(c) displays the Radon transform of the dataset given in Fig. 6(c), where the source is located at position B. Now, there is no symmetry and periodicity anymore, but the basic properties of the preceding figures can be recognized again.

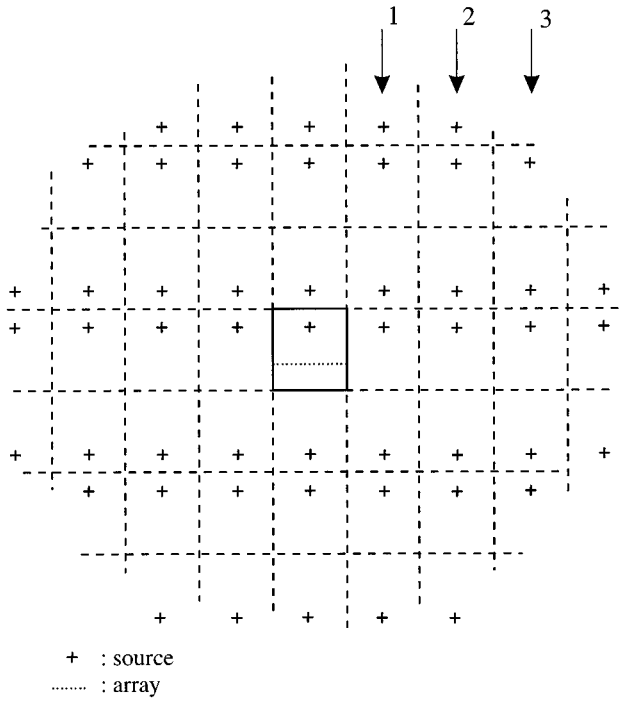


FIG. 8. Two-dimensional mirror-image source model of a rectangular enclosure. Image sources in the columns marked ''1,'' ''2,'' etc. correspond with the data clusters marked identically in Fig. 7(b).

IV. MEASURED RESPONSES OF AN OMNIDIRECTIONAL LOUDSPEAKER IN A LECTURE HALL

In the lecture hall, as described in the Appendix, measurements have been carried out resembling one of the simulations discussed in the previous section—the one with the source positioned symmetrically with respect to the array (source position A). The measurements are shown in Fig. 9(a). For comparison, the simulated data earlier shown in Fig. 6(b) are displayed in Fig. 9(b). It is seen that there is a very good agreement between measured and simulated results. Reflections from the left-hand sidewall are now weaker than those from the opposite side due to a difference in absorption coefficient caused by the presence of windows. An ''extra'' hyperbola can be seen (see arrow) that is caused by a level jump in the floor which is not included in the simulation model. Figure 10 gives the result of the Radon transform of Fig. 9, representing the data in the (p,τ) domain. The measured response of the level jump in the floor is now found [Fig. 10(a)] as additional elliptic curves not present in the simulated data [Fig. 10(b)]. Much more evident than in the (ξ,t) domain are now the effects of diffraction (due to finite boundaries and diffusive ceiling) in the measurements: the well-defined clusters in the simulated data are now blurred in the measurement results. Moreover, in the measured data the clusters at the right-hand side, representing the reflections from the left-hand sidewall, are now weaker than in the simulations, due to the aforementioned difference in absorption coefficient.

V. SPATIAL FILTERING IN THE RAY PARAMETER VERSUS INTERCEPT TIME DOMAIN

The (p,τ) domain is very well suited to select wave field components incident under certain angles or reflected at

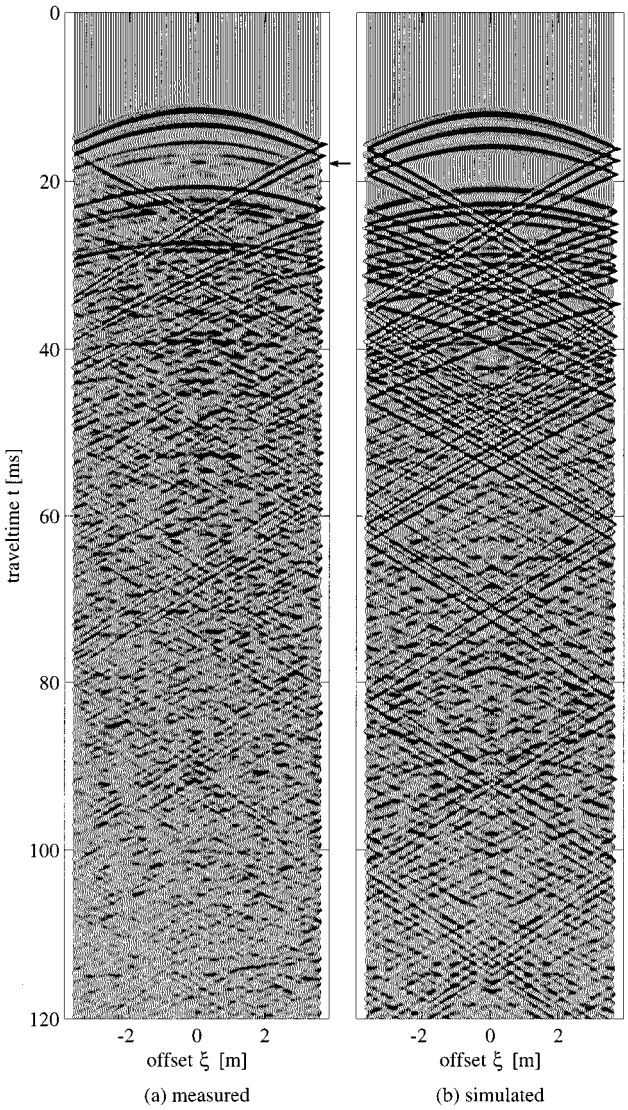


FIG. 9. Point source response in the offset versus travel time domain.

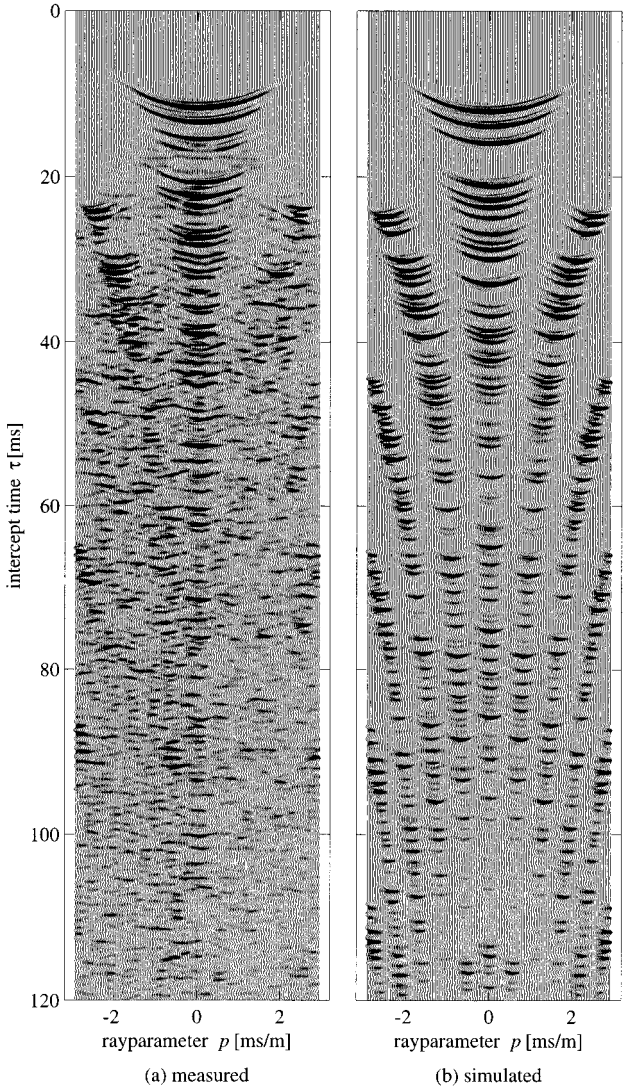


FIG. 10. Point source response in the ray parameter versus intercept time domain.

specific boundaries: Such events are characterized by specific ray parameter and intercept time values and can easily be separated by windowing in the (p,τ) domain. By applying an inverse Radon transform to the modified (p,τ) domain, the selected waves can be represented in the offset versus travel time domain. As an example, the data cluster marked ''1'' in the simulation of Fig. 7(b) and representing all waves reflected once against the right-hand sidewall, has been selected by zeroing all other data; a Hanning window was used to avoid truncation effects. The result is shown in Fig. 11(a). Figure 11(b) displays the data after inverse Radon transformation, showing the selected waves in the offset versus travel time domain. They can be identified as events also present in the full offset-travel time response given in Fig. 6(b). Giving another example, the lateral reflections in Fig. 6(b), i.e., wavefronts incident to the microphone array under angles between 60 and 90 degrees, are selected by zeroing all data beyond the slowness range $|p|<5.9\times10^{-3}$ s/m in Fig. 7(b) [see result in Fig. 11(c)], followed by an inverse Radon transform. The result is displayed in Fig. 11(d). Comparison with Fig. 6(b) shows that, indeed, all nonlateral wavefronts have been filtered out by this operation.

VI. APPLICATION OF RECIPROCITY

So far, we have discussed the response of a single point source recorded by a microphone array. If we interchange the positions of the point source and the n-th microphone, then, according to the principle of reciprocity, the same response is measured at the original source position that was previously measured at the n-th microphone position. Now, let us replace all microphones by point sources and the point source by a microphone. If all point sources at the array simultaneously emit the same signal, thus generating a cylindrical wave parallel to the array, all responses previously measured at the microphone array are superposed at the point receiver (i.e., at the original source position). This superposition means that the individual responses are summed, for all travel times, along horizontal time lines. In Sec. I it was shown that this operation is fully identical to the Radon transform according to Eq. (11a) for $p=0$. We may conclude that the room response of the cylindrical source wave field is given by the trace $p=0$ of the Radon transformed microphone array recording.

If each point source of the array (i.e., the original re-

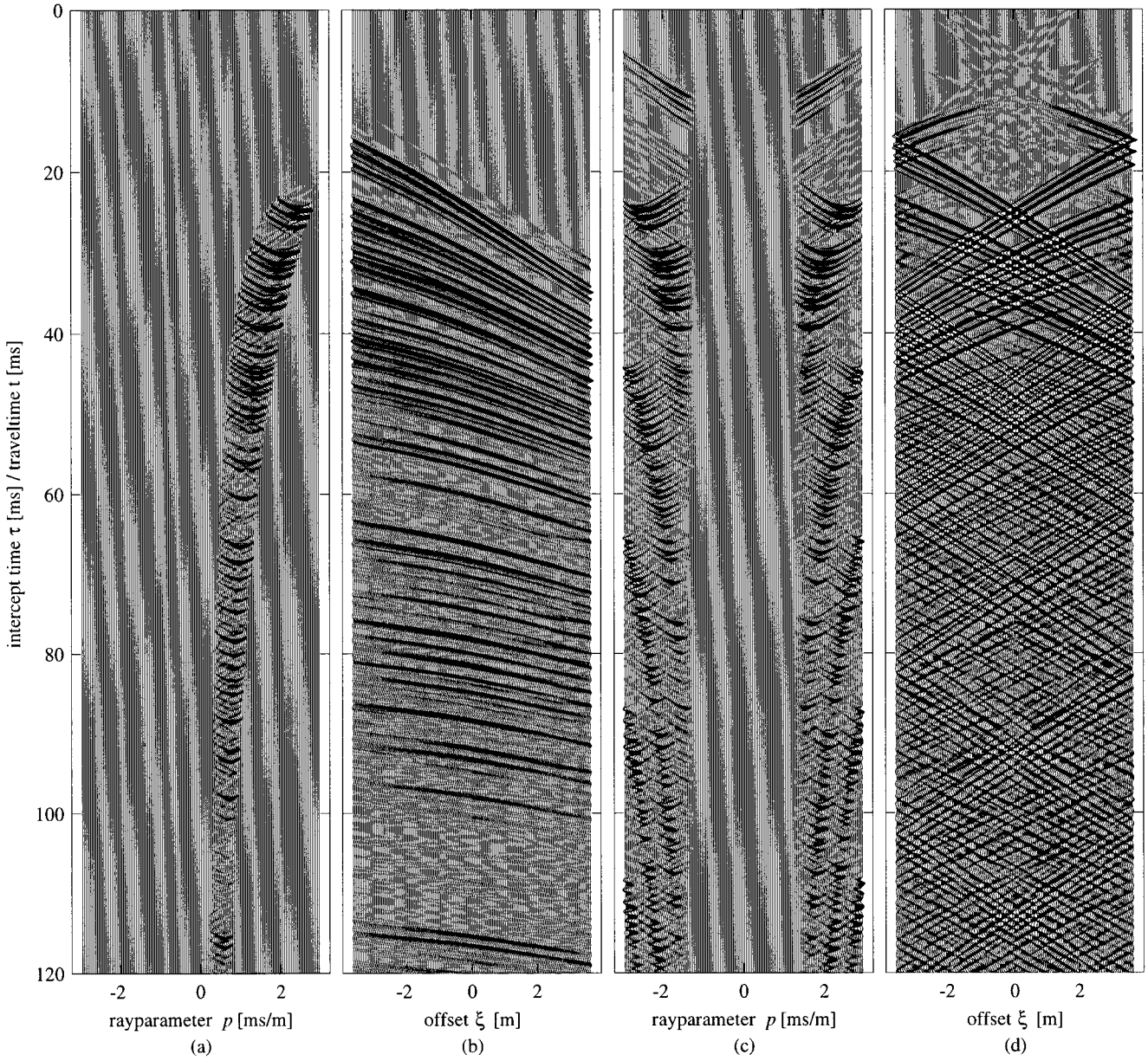


FIG. 11. Spatial windowing in the ray parameter versus intercept time domain and inverse Radon transformation. (a) Selection in the (τ,p) domain of all waves reflected once at the right-hand sidewall; (b) same waves in the (ξ,t) domain after inverse Radon transformation; (c) selection in the (τ,p) domain of all waves with lateral incidence on the array (angle with normal $>\pi/3$); (d) same waves in the (ξ,t) domain.

ceiver array) emits its signal a time Δt later than its neighbor, where

$$\Delta t = \frac{\Delta \xi \tan \alpha}{c}, \tag{20}$$

then, at the point receiver (i.e., at the original source position), all responses originally measured at the microphone array are superposed with a delay Δt between adjacent traces. This is equivalent to summing, for all travel times, the data along a time line making an angle α with the ξ axis, i.e., performing the Radon transform for $p=\sin \alpha/c$. Hence, the room response of this tilted cylindrical source wave field is fully determined by the trace $p=\sin \alpha/c$ of the Radon transformed microphone array recording. This holds for any angle α.

Summarizing, the room response of any prespecified sound field emitted by a point source array can be synthesized at a receiver position by using the principle of reciprocity and measuring the response of a point *source* at the position of this receiver with *microphones* at the array position. Figure 12 shows the results of a synthesized *focused* array, using the microphone array recording of Fig. 6(a). If we choose the focus point at the original source position, then the focused source response can be synthesized by time shifting the original array measurements [Fig. 12(a)] such that the hyperbolic direct wavefront appears as a horizontal plane wavefront [Fig. 12(b)], followed by superposing all the shifted microphone signals to one synthesized response [Fig. 12(c)].

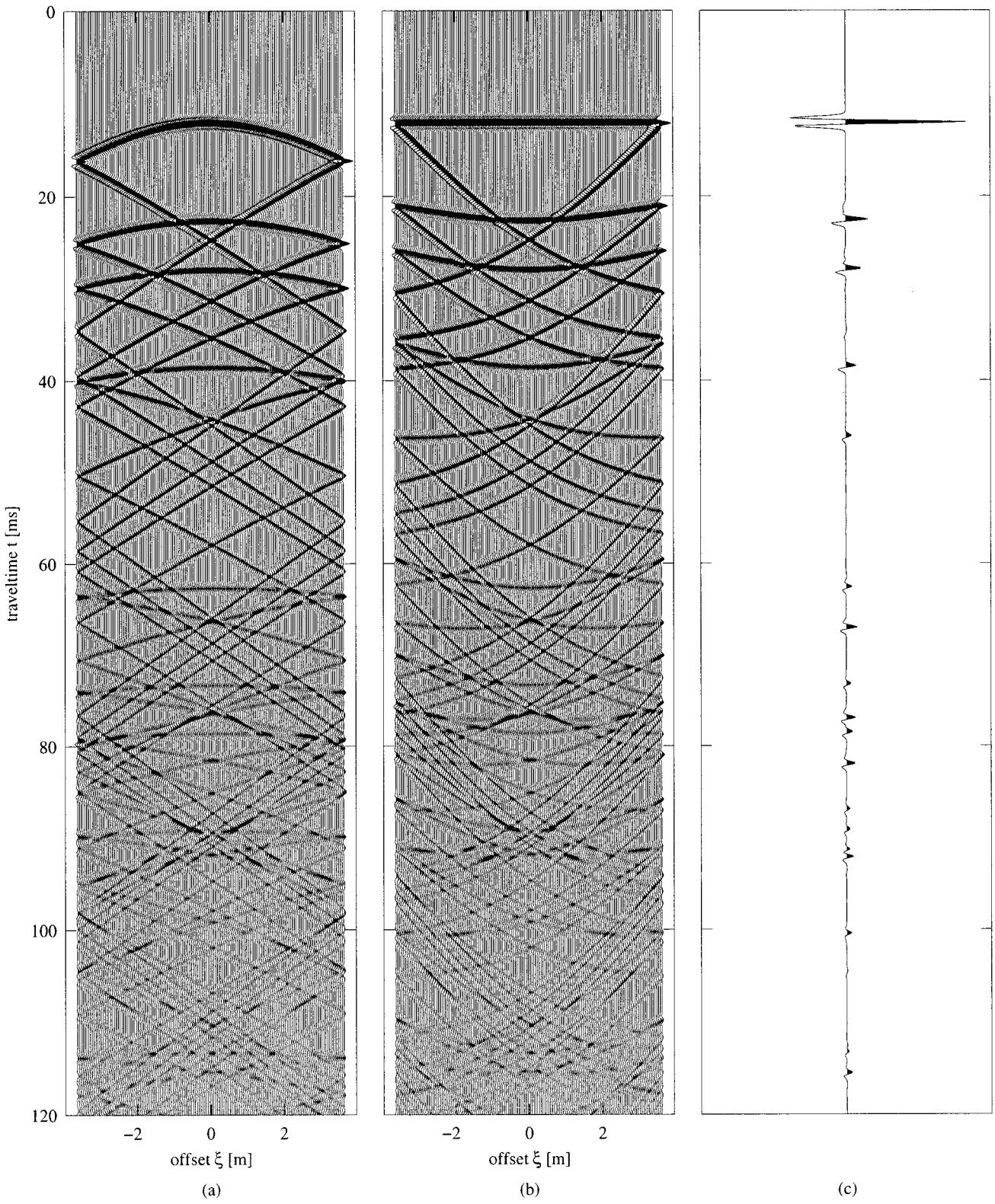


FIG. 12. Synthesis of the response of a focused source array by applying the principle of reciprocity to the microphone array measurement in Fig. 6(a). (a) Microphone array recording (input); (b) result of the time-shifting process; (c) synthesized response (output).

The foregoing also means that by recording the array response (e.g., directed along a horizontal ξ coordinate) of a point source with variable positions (e.g., directed along a vertical η coordinate), array responses are available along two perpendicular directions.

VII. CONCLUSIONS

(1) It has been proposed to make a fundamental change in the way impulse responses are measured and analyzed in enclosed spaces. Instead of using a sparse distribution of

receiver positions with single microphones, arrays of closely spaced microphones are proposed to better register the properties of complex sound fields. Instead of using statistical analysis tools based on signal theory, the resulting multichannel recordings allow the application of deterministic analysis tools based on wave theory.

(2) Numerically simulated measurements as well as real measurements confirm that array measurements facilitate the identification of wavefronts in complex wave fields.

(3) The linear Radon transform has been introduced as a powerful wave-field-oriented analysis method. Application of this transform allows a quantitative study of the directional properties of the wave field under consideration. In addition, the Radon transform allows the decomposition of the total wave field into subfields that can be related to the major reflecting boundaries of the enclosure.

(4) Analysis of array measurements shows that in practical situations boundary diffractions play an important role. Study of these diffraction phenomena may indicate how to refine current numerical modeling algorithms.

(5) Using the principle of reciprocity, microphone array measurements allow the synthesis of responses in enclosed spaces at the source position(s), generated by a synthesized source wave field with a prespecified shape at the microphone array(s). By recording the array responses (e.g., directed along a horizontal ξ coordinate) of a point source with variable positions (e.g., directed along a vertical η coordinate), array responses are available along two perpendicular directions.

It is expected that a better description of complex wave fields in practical situations may generate more insight in the way these wave fields are perceived. The integration of physical acoustics with psychoacoustics is one of the subjects of our current research.

ACKNOWLEDGMENT

The authors gratefully thank Jan Baan for his assistance in generating the simulations.

APPENDIX: DESCRIPTION OF ENCLOSURE AND POSITIONS OF SOURCE AND ARRAY

The room considered in the simulations has a length of 8.61 m, a width of 7.35 m, and a height of 2.92 m; see Fig. A1. The origin of the Cartesian coordinate system is chosen in the center of the room. In the simulations, the absorption coefficients of the floor and the walls have been chosen 0.1 for all boundaries, independent of frequency and angle of incidence. For the ceiling, the value is 0.4. The sound source is a monopole generating a wavelet with bandwidth 500–1300 Hz. One source position (''A'') is located symmetrically with respect to the side walls ($x_s=0$, $y_s=2.5$ m, $z_s=0.34$ m), another (''B'') is chosen at $x_s=2.0$ m and the same values for y_s and z_s. The wave field is recorded with a room-wide microphone array at $y_r=-1.60$ m oriented in parallel with the front and back walls at a height $z_r=-0.26$ m. The interspacing of the omnidirectional array microphones is 0.05 m to avoid spatial aliasing.

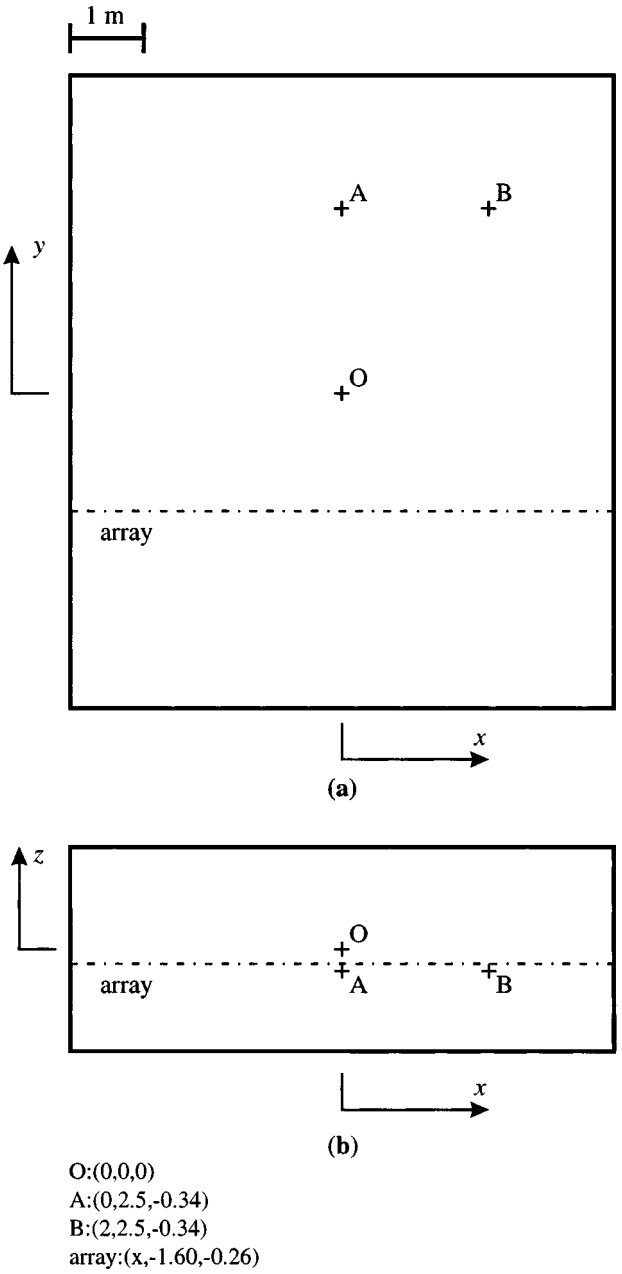


FIG. A1. Geometry of the enclosure with source and microphone array positions. (a) Floorplan; (b) sideview.

Measurements have been carried out in a lecture hall having the same dimensions, with the array at the same position as in the simulation and the source at position A. The directivity patterns of the loudspeaker and the array microphones are approximately omnidirectional within the bandwidth of the signal. For the front, back, and right-hand sidewall, 0.1 is a good approximation of the absorption coefficient value. The same holds for the floor from which the chairs were removed. A raiser with a height of ca. 0.10 m is fixed on the floor extending from $y=-1.90$ m to the back wall; its edge causes diffraction effects that are not included in the simulation. The left-hand sidewall has an absorption coefficient of ca. 0.8. The ceiling has an absorption coefficient of ca. 0.4 and is diffusive.

[1] V. L. Jordan, ''A group of objective acoustical criteria for concert halls,'' Appl. Acoust. **14**, 253–266 (1981).
[2] G. M. Hulbert, D. E. Baxa, and A. Seireg, ''Criterion for quantitative rating and optimum design of concert halls,'' J. Acoust. Soc. Am. **71**, 619–629 (1982).
[3] A. J. Berkhout, M. M. Boone, and P. J. M. Valks, ''Multichannel impulse responses for outdoor sound propagation,'' J. Acoust. Soc. Am. **98**, 1169–1177 (1995).
[4] S. R. Deans, *The Radon Transform and Some of its Applications* (Wiley, New York, 1983).

Effects of air conditioning on sound propagation in a large space

Yasutaka Ueda
Graduate School of Science and Technology, Kobe University, Rokkodai Nada Kobe 657, Japan and Research and Development Center, Hazama Corporation, Karima Tsukuba 305, Japan

Yoichi Ando
Graduate School of Science and Technology, Kobe University, Rokkodai Nada Kobe 657, Japan

(Received 26 July 1996; revised 25 July 1997; accepted 30 July 1997)

Sound-pressure level (SPL) fluctuation caused by air conditioners is measured using pure tones as the sound source, in a gymnasium. When the air conditioner is turned on, the mean standard deviation of SPL changes about 0.9 dB at 2 kHz, 2.5 dB at 4 kHz, 3.3 dB at 8 kHz, and 3.8 dB at 16 kHz, while that in the case with air conditioning off is less than 0.6 dB at all frequencies. The cumulative density of mean-squared amplitude corresponds well with the gamma distribution. Based on this statistical analysis, the SPL change is thought to be related to the impulse response which is represented by the direct sound (regular) and changing delay time of reflection sound (irregular). The cumulative density of mean-squared amplitude for this model using pure tones also indicated a gamma distribution. It is found that the frequency characteristic and the cumulative density of the SPL change have the same tendencies as the method of modulation for the delay time of reflected sound. © *1997 Acoustical Society of America.* [S0001-4966(97)05011-X]

PACS numbers: 43.55.Br, 43.58.Fm [SLE]

INTRODUCTION

More and more buildings with large indoor spaces have been built in recent years. The air conditioning systems used in such buildings have also increased in size, leading to significant changes in the physical environment. There are many studies[1–8] on the effects of air turbulence on sound transmission outdoors. In outdoor sound propagation in air turbulence, sound-pressure level (SPL) fluctuation and reduction are caused by phase changes and the interference of waves. In this paper the effects of air currents produced by an air conditioning system on sound transmission in a gymnasium are investigated.[9] This investigation was done to assess how the dynamic environment, such as air currents produced by an air conditioning system, would influence sound propagation in a large indoor space. It is confirmed that there is a large fluctuation in the SPL for high-frequency ranges. Based on statistical analysis of the SPL fluctuation, a sound propagation model that incorporates a time variant system is proposed.

I. SPL MEASUREMENT

SPL was measured in a gymnasium because the gymnasium allows for a large mean-free path, and its interior is made with hard materials which lower the attenuation of reflected sounds (1:89.4 m w: 67.6 m h: 18.0 m 5000 seats and RT: 2.1 s at 500 Hz). Table I shows the experimental conditions. The maximum air speed at the outlet duct was about 8 m/s. We could not measure the air speed and the direction throughout the room, but the air speed did not exceed 0.5 m/s in the audience area at 1.5 m above the floor level (case 2). There are 36 outlet ducts with a diameter of 450 mm and 44 ducts with a diameter of 750 mm on the ceiling. The vertical temperature difference from floor to ceiling did not exceed 3 °C in the condition of the air conditioner on or off. Temperature was uniform in the horizontal plane. A dodecahedron loudspeaker was used to produce pure tones of 0.5 and 1 kHz, and a plasma loudspeaker was used for 2-, 4-, 8-, and 16-kHz frequencies. Figure 1 shows the location of the sound source and observation points. The height of the sound source was 1.5 m above the floor, and observation points were 1.2 m above the floor. The distance between the sound source and the observation points was from 27 m to 57 m. The SPL was recorded for about 1 min at each observation point.

II. RESULTS OF MEASUREMENT

A. Fluctuation of SPL

Figure 2 shows the SPL fluctuation recorded at observation point $K1$ (0.5–16 kHz). Figure 3 shows the mean standard deviation of SPL which was measured at 11 observation points. In case 1, the strong SPL change was not found at any frequency or observation points. In case 2, strong SPL fluctuations were observed at frequencies higher than 2 kHz. As the frequency increases, SPL fluctuation became stronger. These tendencies were apparent regardless of the distance between the sound source and the observation point.

TABLE I. Experimental conditions.

Case	Air conditioner	Air speed in audience area [m/s]	Noise criteria
1	off	≈ 0.0	25
2	on	< 0.5	40

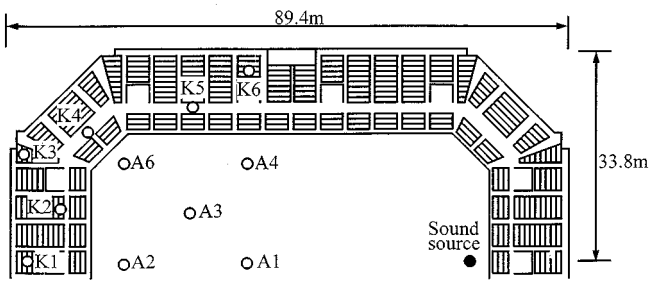


FIG. 1. Location of observation points. ●: sound source, ○: observation points.

B. Statistical analysis of amplitude

When waves of differing amplitude and phase are synthesized irregularly, they can be characterized by a Rayleigh or Nakagami–Rice distribution.[10–13] The measured amplitude was compared with a gamma distribution which includes Rayleigh, Gauss, and approximate Nakagami–Rice distributions, to evaluate how reflected sounds are composed. The general formula of this distribution is

$$p(x;P^2)=\frac{1}{m^{l+1}\Gamma(l+1)}\,x^l e^{-x^2/m}, \tag{1}$$

where $\overline{p^2}[=m(l+1)]$ is the mean value and $\sigma(=m\sqrt{l+1})$ is the standard deviation.[11] The parameter l represents the degree of fluctuation. When l equals 1.0, this distribution is equivalent to a Rayleigh distribution, and when $l>1.0$, it is approximately equal to a Nakagami–Rice distribution. Since the simple correlation coefficient between the distribution of measured data and theory is more than 0.98, we chose to use the values of l and m by changing the parameter l gradually such that $l=0,1,2,\ldots$. Figure 4 shows an example of the cu-

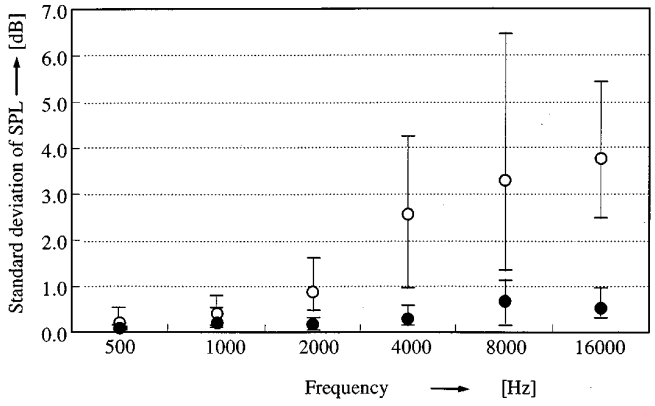


FIG. 3. Mean standard deviation of SPL. ● in case 1 and ○ in case 2. Top and bottom bars show the maximum and minimum standard deviation of SPL.

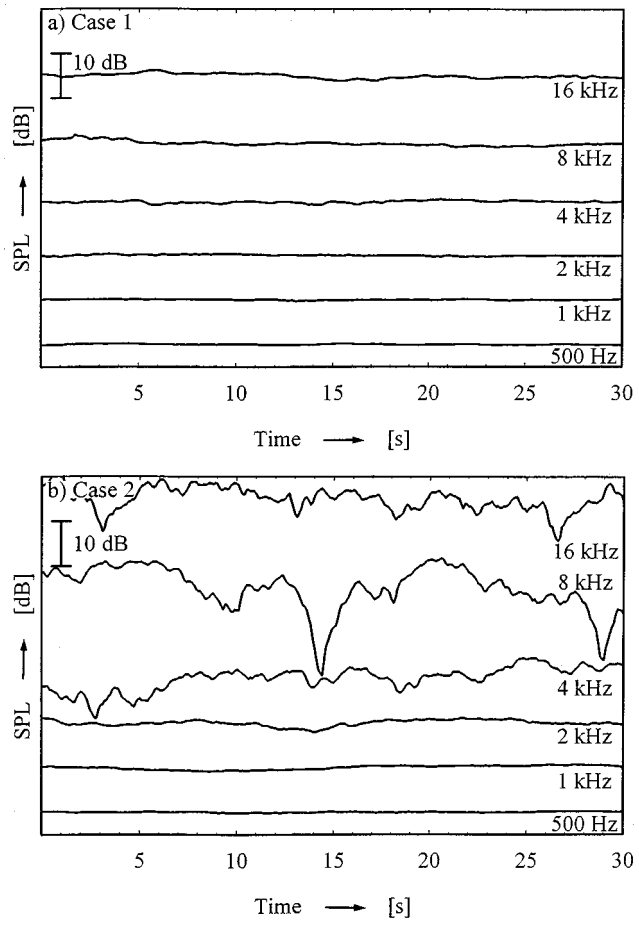


FIG. 2. Sound pressure level at observation point $K1$. (a) in case 1 and (b) in case 2.

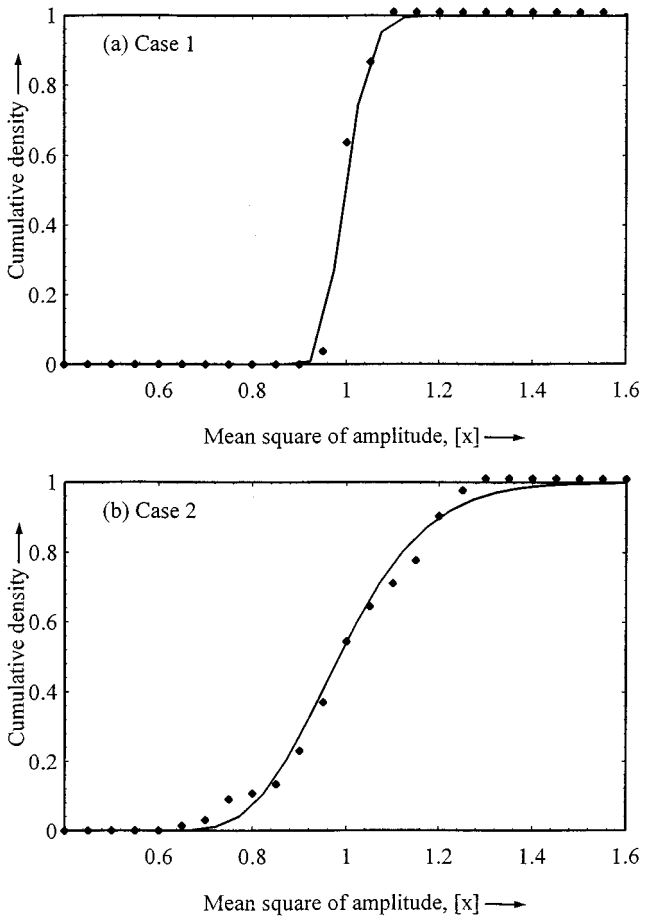


FIG. 4. Cumulative density of the mean-squared pressure. Observation point: $K1$ Frequency: 2 kHz, (a) case 1; (b): case 2.

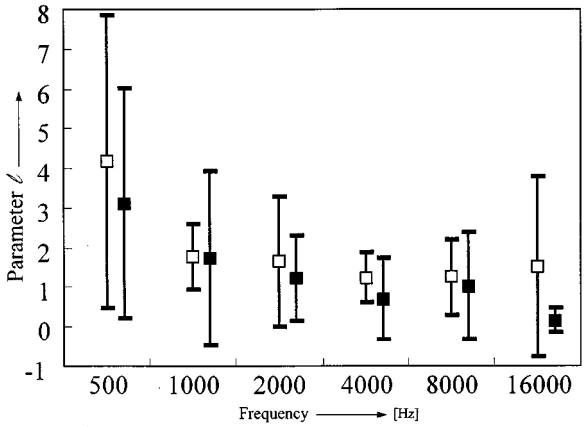


FIG. 5. Mean value of l at each frequency. □ is the mean value of 1 in case 1 and ■ is in case 2. The top and bottom bars show the ±1 standard deviation.

mulative density of the measured values compared with this gamma distribution. Figure 5 shows the mean values of parameter l at each point. In both cases, the measured values agreed with this theoretical curve. The slope of the cumulative density curve for case 1 is steeper than that for case 2. Because l is more than 0.0, the fluctuation of SPL must be caused by the composition of regular waves with irregular waves. As the frequency becomes higher, the value of l becomes smaller. It is considered that the ratio of irregular components to the regular components differed in each case (see Table II).

III. TIME-VARIANT MODEL OF IMPULSE RESPONSE

A. Concept of the time-variant impulse response

The sound pressure at the observation point is usually represented by

$$p(t)=\sum_{n=0}^{\infty} f(t)*R_n h_n(t-t_n), \tag{2}$$

where R_n is the amplitude of each reflection ($n=0$ represents the direct sound) and t_n is the delay time of each reflection.[14] The mechanism of SPL change is represented by the time-variant impulse response model in which the delay time of reflection varies:

$$p(t)=\sum_{n=0}^{\infty} f(t)*R_n h_n(t-t_n(t)). \tag{3}$$

TABLE II. Relationship between statistical characteristics and the condition of wave composition.

Wave composition	Distribution of amplitude	Distribution of mean-squared amplitude
(a) Differing amplitude and phase waves synthesized	Rayleigh distribution	Gamma distribution ($l=0.0$)
(b) Steady waves synthesized with above irregular waves	Nakagami–Rician distribution	Gamma distribution ($l>0.0$)

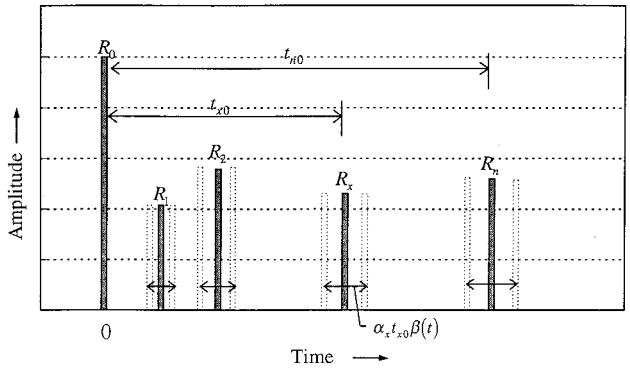


FIG. 6. A time-variant model of impulse response.

In addition, $t_n(t)$ is defined by α_n and $\beta_n(t)$. The representing of the delay time of reflections is introduced as

$$t_n(t)=t_{n0}+\alpha_n t_{n0}\beta(t)=t_{n0}(l+\alpha_n\beta_n(t)), \tag{4}$$

where α_n is the change width of delay time, $\beta_n(t)$ is change the property of delay time, t_{n0} is the initial value for delay time under the geometry ray theory, and $\langle t_n(t)\rangle=t_{n0}$. By substituting the above expression into formula (2), expression (5) is obtained:

$$p(t)=\sum_{n=0}^{\infty} f(t)*R_n h_n(t-t_{n0}(l+\alpha_n\beta_n(t))). \tag{5}$$

It is simply supposed that α_n becomes larger as delay time increases. Figure 6 shows the concept of this time-variant impulse response model.

B. Statistical character

The statistical character of mean-squared pressure for this model was examined as follows. When $f(t)$ is equivalent to sine wave, the sound pressure at a point can be represented

$$\begin{aligned} p(t)&=\sum_n \sin\omega t*R_n h_n(t-t_n)\\ &=\sum_n R_n \sin(\omega t-\omega t_n)\\ &=\{R_0\sin\omega t_0+R_1\sin(\omega t-\omega t_1)+R_2\sin(\omega t-\omega t_2)\\ &\quad+\cdots\}\\ &=\sin\omega t(R_0+R_1\cos\omega t_1+R_2\cos\omega t_2+\cdots)\\ &\quad+\cos\omega t(R_1\sin\omega t_1+R_2\sin\omega t_2+\cdots). \end{aligned} \tag{6}$$

When t_n distributes according to formula (4), ωt_n are simply denoted by a_{ne}^*

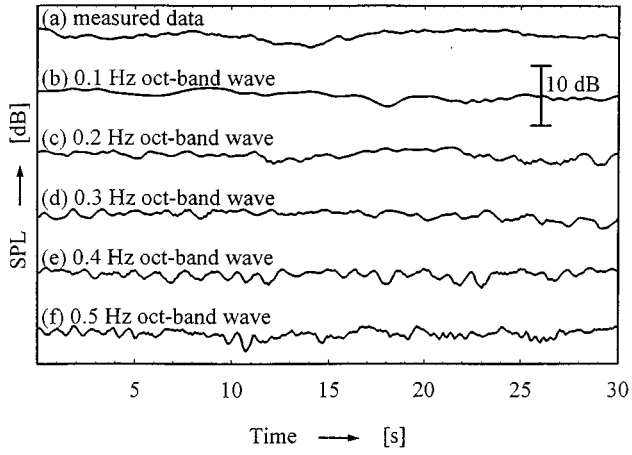


FIG. 7. Comparison of the SPL simulated by some modulation method of delay time. $K1$, 2 kHz, (a) Measured data; (b) by 0.1-Hz octave band; (c) by 0.2-Hz octave band; (d) by 0.3-Hz octave band; (e) by 0.4-Hz octave band; (f) by 0.5-Hz octave band.

$$p(t)=\sin\ \omega t\left(R_0+R_1\sum_i^{nl}\cos\ a_i+R_2\sum_j^{n2}\cos\ a_j+\cdots\right)$$
$$+\cos\ \omega t\left(R_1\sum_i^{nl}\sin\ a_i+R_2\sum_j^{n2}\sin\ a_j+\cdots\right)$$
$$=\sin\ \omega t(R_0+r_1+r_2+\cdots)+\cos\ \omega t(s_1+s_2+\cdots)$$
$$=\sin\ \omega t(R_0+r)+\cos\ \omega t(s). \tag{7}$$

The distribution of a_i is flat ($0\leqslant a_i\leqslant 2\pi$). However, it approaches normality as $n1,n2,\ldots\rightarrow\infty$, due to the central limit theorem. Consequently the distribution of each term including the trigonometric function becomes a normal distribution. The variables r_1,r_2,s_1,s_2, etc., are normally distributed variables. Since the normal distribution reproduces itself by composition, r and s are also normal variables. The fundamental equation of Nakagami–Rice distribution is

$$\overline{p^2}=\tfrac{1}{2}\left[(R_0+r)^2+s^2\right]. \tag{8}$$

Consequently, the distribution of this model can be applied to a gamma distribution approximately.

IV. SOUND PRESSURE LEVEL (SPL) SIMULATION

SPL was simulated to confirm the appropriateness of this time-variant model for the measured SPL change. The initial values for delay time t_{n0} and amplitude R_n of each reflected sound were calculated at each observation point in the gymnasium using the geometrical sound theory. The calculation conditions were that the delay time of reflections was within 500 ms, and the reflection frequency was within three times.

A. Examination of $\beta(t)$

Delay time change $[\beta(t)]$ can be modulated by random noise, periodical waves (like sine waves), and band waves. It can be considered that SPL calculated by random noise would change randomly, and that by periodical waves would change periodically. And there are also many types of modulation methods in band waves. Here, octave-band noise was chosen and examined. The fluctuation width (α_n) was controlled so that it has the same standard deviation as that of the measured data. Figure 7 shows the measured SPL and the calculated SPL modulated by octave-band waves from 0.1 to 0.5 Hz (at observation point $K1$, 2 kHz). As the frequency of the modulation became higher, the simulated SPL fluctuated. Next the frequency envelope of SPL change was examined (Fig. 8). The frequency envelope of simulated SPL change with a 0.2-Hz octave-band wave, resembled that of measured data at this point. Figure 9 shows the SPL from 0.5 to 16 kHz, calculated by a 0.2-Hz octave-band wave.

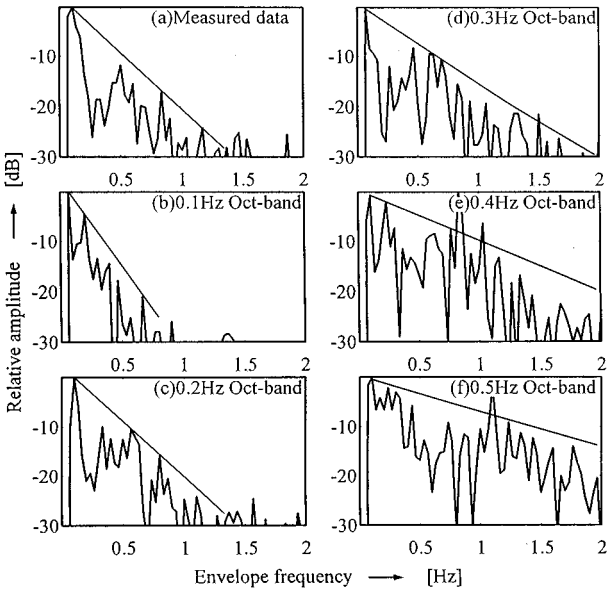


FIG. 8. Comparison of the envelope frequency in SPL change. $K1$, 2 kHz, (a) Measured data; (b) by 0.1-Hz octave band; (c) by 0.2-Hz octave band; (d) by 0.3-Hz octave band; (e) by 0.4-Hz octave band; (f) by 0.5-Hz octave band.

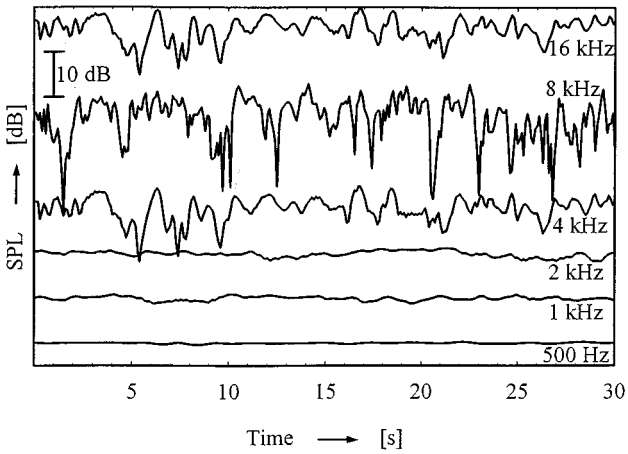


FIG. 9. Calculated SPL by using 0.2-Hz octave-band noise at receiving point $K1$.

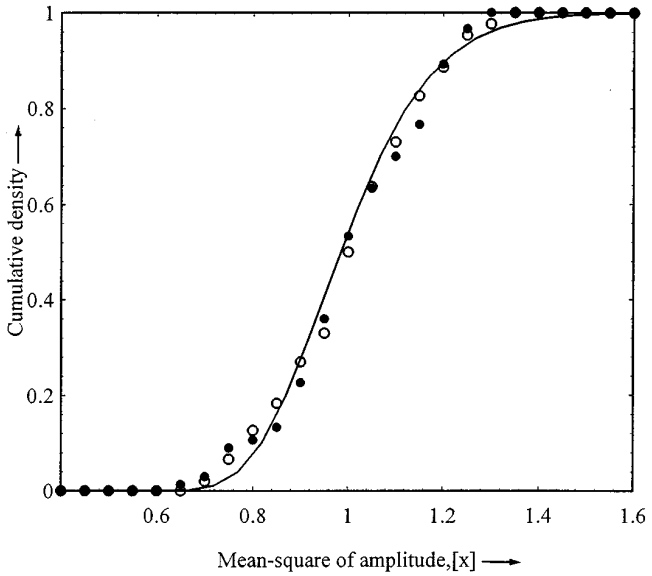


FIG. 10. Cumulative density between the measured and calculated mean-squared amplitude with the gamma distribution. *K*1, 2 kHz. ○: calculated mean-squared pressure; ●: measured mean-squared pressure, the solid line indicates the gamma distribution.

B. Statistical distribution of simulated SPL

The cumulative density of the mean-squared amplitude of Fig. 9 was compared with the gamma distribution of Fig. 10. The cumulative density of the amplitude fits the gamma distribution well, as does the measured data.

V. CONCLUSION

Analyses of SPL and the statistical distribution of amplitude for both measurement conditions (air conditioning on and off) gave the following conclusions.

A change in SPL was observed when the air conditioning was on. The influence is significant in the frequency domain, above 2 kHz. As the frequency becomes higher, the SPL fluctuates more. Since the distribution of amplitude agrees with the gamma distribution, SPL change occurs as a result of the combination of the direct wave (regular) and the changing delay time of reflected sound (irregular).

A time-variant impulse response in which the delay time of reflected sound changes over time was devised. Based on the comparison of the simulated SPL, using some modulation of delay time, with the measured SPL, it is found that SPL fluctuates more as the modulation frequency becomes higher. From the analysis of envelope frequency in the SPL change, the envelope of the measured data resembles that of the 0.2-Hz octave-band wave. The SPL that was calculated by using 0.2-Hz octave-band noise shows the same characteristics as the measured data, and the cumulative density of the mean-squared pressure corresponded to the gamma distribution.

It is almost possible to confirm the appropriateness of the time-variant model, but there is need for more physical analysis. The authors also are proceeding with psychological experiments that use this time-variant model.

[1] G. A. Daigle, ''Correlation of the phase and amplitude fluctuation between direct and ground-reflected sound,'' J. Acoust. Soc. Am. **68**, 297–302 (1980).
[2] U. Ingard and G. C. Maling, Jr., ''On the effect of atmospheric turbulence on sound propagation over ground,'' J. Acoust. Soc. Am. **35**, 1056 (1963).
[3] G. A. Daigle, J. E. Piercy, and T. F. W. Embleton, ''Effect of atmospheric turbulence on the interference of sound wave near a hard boundary,'' J. Acoust. Soc. Am. **64**, 622–630 (1978).
[4] R. J. Thompson, ''Ray theory for an inhomogeneous moving medium,'' J. Acoust. Soc. Am. **51**, 1675 (1972).
[5] R. J. Thompson, ''Ray-acoustic intensity in a moving medium 1,'' J. Acoust. Soc. Am. **55**, 729 (1974).
[6] R. J. Thompson, ''Ray-acoustic intensity in a moving medium 2, stratified medium,'' J. Acoust. Soc. Am. **55**, 733 (1974).
[7] H. Tachibana and K. Ishii, ''Scale model experiment of the effects of wind on sound propagation,'' Inter-Noise **75**, 623 (1975).
[8] L. L. Beranek, *Noise Reduction* (McGraw-Hill, New York, 1960), p. 185.
[9] Y. Ueda, S. Iwashita, and Y. Ando, ''Sound fluctuation due to the air turbulence in a large indoor space,'' Proc. ICA 14, f8-3 (1992).
[10] R. V. Waterhouse, ''Statistical properties of reverberant sound fields,'' J. Acoust. Soc. Am. **43**, 1436–1444 (1968).
[11] L. Maisel, *Probability, Statistics and Random Processes* (Simon and Schuster, New York, 1971).
[12] N. Mohanty, *Random Signals Estimation and Identification Analysis and Applications* (Van Nostrand Reinhold, New York, 1986), pp. 139–141.
[13] H. Cramer, *Mathematical Methods of Statistics* (Princeton U.P., Princeton, NJ, 1946), pp. 126, 233–236.
[14] Y. Ando, *Concert Hall Acoustic* (Springer-Verlag, Berlin, 1985).

Water as a standard in the measurements of speed of sound in liquids

Wojciech Marczak
University of Silesia, Institute of Chemistry, Szkolna 9, 40-006 Katowice, Poland

(Received 29 April 1996; accepted for publication 23 July 1997)

The speeds of sound in water measured by Del Grosso and Mader [J. Acoust. Soc. Am. **52**, 1442–1446 (1972)], Kroebel and Mahrt [Acustica **35**, 154–164 (1976)], and Fujii and Masui [J. Acoust. Soc. Am. **93**, 276–282 (1993)] were compared. A fairly good agreement was found. A new fifth-order polynomial describing the dependence of the speed of sound in water on temperature (ITS-90) within the limits 0–95 °C was proposed. The importance of the effect of impurities of water on the systematic errors of the measurement results was pointed out. The accuracy of the relative measurements of speed of sound in liquids was discussed in the context of the proper choice of a standard (''true'') value of the speed in water. © *1997 Acoustical Society of America.* [S0001-4966(97)01511-7]

PACS numbers: 43.58.Dj, 43.30.Es, 62.60.+v [SLE]

INTRODUCTION

The accuracy of the relative measurements of the speed of sound in liquids is determined by the applied velocimeter and the assumed ''true value'' of the speed of sound in the standard liquid applied for calibration, i.e., most commonly in water. For the absolute speed measurements the problem of a reliable standard seems to also be of great importance; in those experiments the standard should assure the simplest method of checking the measuring device.

Contrary to the density, heat capacity, and many other physicochemical quantities, the critically evaluated values, which have been published in numerous monographs and tables, high precision data of the speed of sound in water are available only in original papers.[1–7] Alternatively, the speed of sound in water can be calculated using the equation of state proposed by the International Association for the Properties of Steam (IAPS);[8] the procedure, however, is labor consuming and leads to results of insufficient accuracy.

The speed of sound in water as a function of temperature has been extensively measured by the application of a variety of methods. Probably, the most nearly accurate results are those of Del Grosso and Mader,[1] Kroebel and Mahrt,[2] and Fujii and Masui.[3] The other ones are either too scattered,[4] or even systematically shifted,[5–7] as has been discussed by Fujii and Masui.[3] In this paper the various speeds of sound in water under atmospheric pressure given by Del Grosso and Mader,[1] Kroebel and Mahrt,[2] and Fujii and Masui[3] have been compared and, in order to get reliable standard values of the speed of sound in water, their temperature dependence was described by a single fifth-order polynomial.

I. COMPARISON AND DISCUSSION OF THE DATA

Some details of the measurement techniques of the speed of sound in pure water are listed in Table I. In order to ensure the consistency of all the calculations, the International Temperature Scale of 1990 (ITS-90)[9] was applied in further considerations. Therefore the temperatures reported by Del Grosso and Mader[1] and Kroebel and Mahrt[2] had to be recalculated.

As shown in Fig. 1, there are few experimental points in some temperature intervals. Between 10 and 20 °C the speed of sound was measured only by Kroebel and Mahrt.[2] From 35 °C to 75 °C the speeds were given at 5-deg intervals by Fujii and Masui[3] and also, but rarely, by Del Grosso and Mader.[1] Within the limits 75–100 °C only five speed values, two at 90 °C and three at 95 °C, measured by Del Grosso and Mader[1] are available.

In the papers cited,[1–3] the dependencies of the speed of sound (c) on temperature (t) were approximated by fifth-order polynomials with the standard deviations of 0.003, 0.018, and 0.0024 m s^{-1}, respectively. Del Grosso and Mader[1] have reported the coefficients of polynomial. Kroebel and Mahrt[2] published a corrected version of the polynomial of Del Grosso and Mader[1] with new values of the intercept and the linear coefficient. Moreover, the results of Del Grosso and Mader[1] were recalculated from the International Practical Temperature Scale of 1968 (IPTS-68) into the ITS-90 by Bilaniuk and Wong,[10] who proposed a new fifth-order polynomial approximating the dependence of c on t_{90} (where t_{90} is the temperature in °C according to the ITS-90).

In order to compare the results of Del Grosso and Mader,[1] Kroebel and Mahrt,[2] and Fujii and Masui,[3] a fifth-order polynomial was fitted to all the 209 experimental points given by these authors:

$$c = \sum_{i=0}^{5} a_i t_{90}^i . \tag{1}$$

The coefficients a_i found by the least squares method are given in Table II, while the residuals are plotted in Fig. 2. The mean deviation from the regression line equals 0.012 m s^{-1} and is of the same order of magnitude as the declared accuracies of the measurements (Table I). As shown in Fig. 2, the data obtained by Del Grosso and Mader[1] and Fujii and Masui[3] seem to be more nearly precise than those

TABLE I. The comparison of the measurements of the speed of sound in water.

Authors	Method of measurement	Ultrasonic frequency (MHz)	Accuracy ($m\,s^{-1}$)	No. of experimental points	Temperature range (°C)
Del Grosso and Mader[a]	Interferometer	5	0.015	148	0–95
Kroebel and Mahrt[b]	Interferometer	4	0.04	20	3–34
Fujii and Masui[c]	Coherent phase detection	16	0.013	41	20–75

[a]See Ref. 1.
[b]See Ref. 2.
[c]See Ref. 3.

of Kroebel and Mahrt.[2] It should be noted that the observed systematic variation of the deviations with temperature cannot be due to the uncertainty in the temperatures that have been declared to be 0.001 °C by Del Grosso and Mader,[1] 0.005 °C by Kroebel and Mahrt,[2] and 0.003 °C by Fujii and Masui.[3] The differences between the speeds of sound determined experimentally by Fujii and Masui[3] and those calculated from the polynomial of Bilaniuk and Wong[10] (based on the data of Del Grosso and Mader[1]) are plotted against temperature in Fig. 3, together with the error lines for a temperature deviation of 0.01 °C. Thus even assuming a higher tolerance limit for the temperature than those declared by the authors cited,[1–3] it is rather impossible, or at least difficult, to explain the observed differences between reported speeds of sound. The trial to introduce a systematic correction to the temperatures of measurements given by Del Grosso and Mader[1] or by Fujii and Masui[3] does not improve significantly the consistency of the results.

It should be noted, however, that the declared uncertainties in temperature may be underestimated. As results from an earlier study performed in our laboratory,[11] the temperature inhomogeneities, measured in a liquid sample of ca. 80-cm^3 volume, can sometimes reach 0.01 °C, while the temperature fluctuations at a given point during a typical speed of sound measurement are approximately ten times smaller. Kroebel and Mahrt[2] have found that the temperature problem was the main source of error in their experiment and estimated the error to be smaller than 0.04 m/s. The total volume of the sample used by Kroebel and Mahrt was 250 dm^3; both of the two other sample cells[1,3] were essentially smaller. Fujii and Masui,[3] who measured the differences in temperatures of various parts of their apparatus by a thermopile, declared that there was no serious problem arose from the temperature inhomogeneities.

According to the theoretical principles of the classical calculus of errors, the errors originating from different sources are assumed to add up. In practice, the errors rarely combine and propagate in such a biased way since some errors of opposite sign cancel each other.[12] Thus the errors of speed of sound, calculated in these papers[1–3] using the assumption of additive errors, are expected to be close to the possible upper limit of the absolute error. The real accuracies of the measurements are probably considerably better than those declared by the authors,[1–3] and a better consistency of the results should be expected. It is reasonable to search for a source (or sources) of error other than those taken into account by the authors.[1–3]

A probable source of systematic errors, apart from those due to the apparatus, seems to be the impurities in the investigated samples of water. Unfortunately, the authors cited[1–3] have not specified the purity of samples used; even the electrolytic conductivities have not been reported. According to Kell, who discussed in detail the influence of dissolved gases and the isotopic composition on the density of water,[13] an increase of 1.5 p.p.m. in density (caused by the presence of

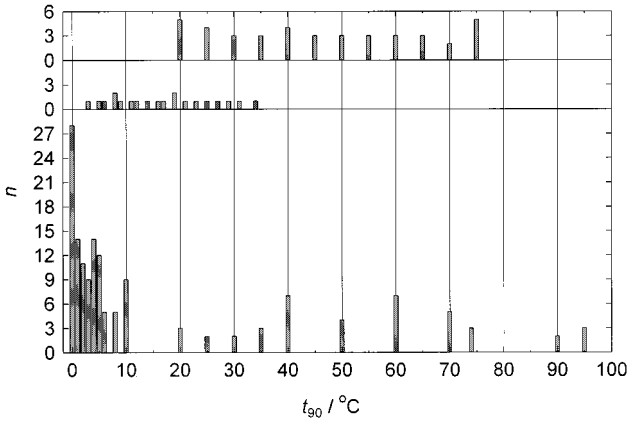


FIG. 1. Histogram of the number of experimental points (n) within 1-deg temperature intervals, reported by Del Grosso and Mader (Ref. 1, bottom), Kroebel and Mahrt (Ref. 2, middle), and Fujii and Masui (Ref. 3, top).

TABLE II. The coefficients a_i of the fifth-order polynomial [Eq. (1)], approximating the dependence of the speed of sound in water, c ($m\,s^{-1}$), on temperature, t_{90} (°C), fitted to 209 experimental values of Del Grosso and Mader,[a] Kroebel and Mahrt,[b] and Fujii and Masui.[c]

i	a_i
0	$1.402\,385\times10^{3}$
1	$5.038\,813\times10^{0}$
2	$-5.799\,136\times10^{-2}$
3	$3.287\,156\times10^{-4}$
4	$-1.398\,845\times10^{-6}$
5	$2.787\,860\times10^{-9}$

[a]See Ref. 1.
[b]See Ref. 2.
[c]See Ref. 3.

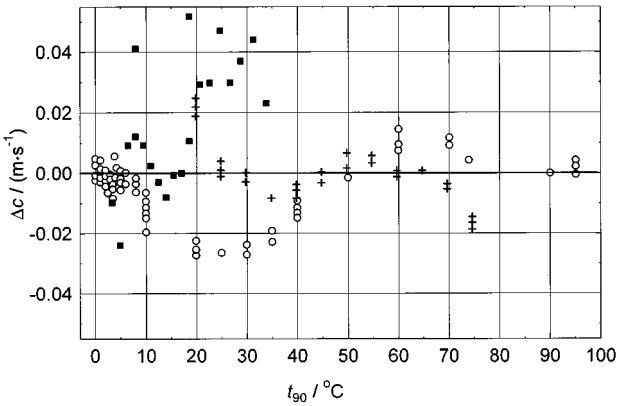


FIG. 2. The residual errors (Δc) of the fit of the fifth-order polynomial [Eq. (1)] to the data of Del Grosso and Mader (Ref. 1, ○), Kroebel and Mahrt (Ref. 2, ■), and Fujii and Masui (Ref. 3, +); $\Delta c = c$(experimental) $- c$(calculated).

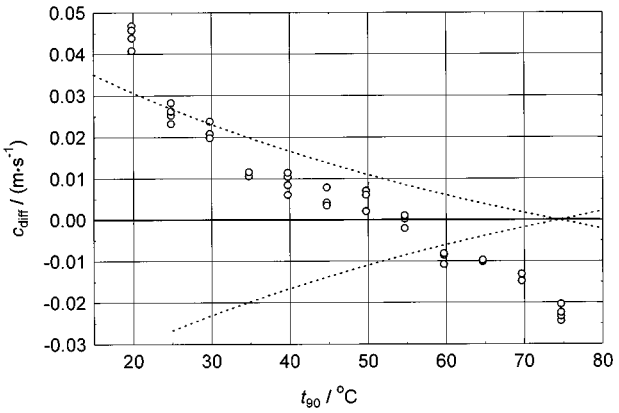


FIG. 3. The difference $c_{\mathrm{diff}} = c_{\mathrm{FM}} - c_{\mathrm{BW}}$ between the speeds of sound measured by Fujii and Masui (Ref. 3), c_{FM}, and those calculated from the polynomial of Bilaniuk and Wong (Ref. 10), c_{BW} (points); dotted lines mark the area of the possible discrepancies in speed of sound caused by a temperature error of 0.01 °C.

deuterium oxide) results in an increase of 1 p.p.m. in the speed of sound. Christiansen *et al.*[14] stated that the difference in density of different fractions of distilled water, due to variations in the isotopic composition, can reach 20 p.p.m. Thus it could be expected that the speed of sound in water can vary, depending on its isotopic composition,[13] up to 13 p.p.m. i.e., 0.02 m s^{-1}. One can assume that dissolved air may cause similar changes in the speed of sound.

II. CHOICE OF A ''TRUE VALUE'' OF SPEED OF SOUND

For accurate measurements of the speed of sound in liquids, when the expected absolute error should not exceed a few centimeters per second, water may be considered to be a reliable standard provided that proper (''true'') values of speed are applied. In the author's opinion, if the temperature is expressed in ITS-90, the most suitable are the speeds calculated from polynomial (1), or those given by Fujii and Masui[3] or Bilaniuk and Wong.[10] Obviously, to achieve such accuracy, the same temperature scale must be applied in all the measurements and calculations.

In the measurements of the speed of sound with an accuracy of the order of a few decimeters per second, any of the speeds given in sources cited,[1–3,10] or in this work, either the measured ones, or those calculated from the appropriate polynomial, can be used as reference values. For such an accuracy the discrepancy in speeds caused by the difference in temperature scales, ITS-90 (t_{90}) and IPTS-68 (t_{68}) is negligible. As calculated by Bilaniuk and Wong,[10] the difference in the absolute values of the speed of sound caused by the shift of the temperature reaches 0.022 m s^{-1} at 100 °C and is smaller for $t < 100$ °C going through a local maximum at ca. 35 °C. For typical organic liquids, an increase in temperature of 1 °C results in a decrease of the speed equal to ca. 4.5 m s^{-1}. At 100 °C, where $t_{68} - t_{90} \approx 0.026$ °C, the discrepancy in speed equals ca. 0.12 m s^{-1}. Thus at 100 °C the total error of the speed of sound in most of the organic liquids, measured by a velocimeter calibrated with water, should not exceed ca. 0.15 m s^{-1}, and ought to be smaller at lower temperatures.

III. CONCLUSIONS

A comparison and the discussion of the experimental data[1–3] lead to the following conclusions:

(1) The speeds of sound in water measured by Del Grosso and Mader,[1] Kroebel and Mahrt,[2] and Fujii and Masui[3] are sufficiently consistent within the limits of the accuracies claimed by these authors.
(2) The adjustments introduced to the polynomial of Del Grosso and Mader[1] by Kroebel and Mahrt[2] seem to be unnecessary.
(3) It seems to be justified to apply the polynomial (1), with the coefficients given in Table II for the calculation of the speed of sound in pure water over the t_{90} range 0–95 °C. As per the results from the scattering of the points plotted in Fig. 2, the error of the calculated speeds probably does not exceed 0.03 m s^{-1}, except for the temperature range 75–90 °C, where the lack of experimental points makes such an estimation impossible.
(4) For measurements of the speed of sound in liquids with an accuracy less than a few decimeters per second any speed given in the literature,[1–3,10] or calculated according to Eq. (1), can be applied as a reference value, independent of the temperature scale (ITS-90 or IPTS-68) used.
(5) In further investigations of the speed of sound in water, the emphasis should be put on its quality; the content of dissolved air and the isotopic composition of water ought to be defined. The use of Standard Mean Ocean Water[15] should be considered.

ACKNOWLEDGMENT

The author is profoundly indebted to Professor S. Ernst for fruitful discussions and critical reading of the manuscript.

[1] V. A. Del Grosso and C. W. Mader, ''Speed of sound in pure water,'' J. Acoust. Soc. Am. **52**, 1442–1446 (1972).
[2] W. Kroebel and K.-H. Mahrt, ''Recent results of absolute sound velocity measurements in pure water and sea water at atmospheric pressure,'' Acustica **35**, 154–164 (1976).

[3] K. Fujii and R. Masui, ''Accurate measurements of the sound velocity in pure water by combining a coherent phase-detection technique and a variable path-length interferometer,'' J. Acoust. Soc. Am. **93**, 276–282 (1993).
[4] H. J. McSkimin, ''Velocity of sound in distilled water for the temperature range 20–75 °C,'' J. Acoust. Soc. Am. **37**, 325–328 (1965).
[5] M. Greenspan and C. E. Tschiegg, ''Speed of sound in water by a direct method,'' J. Res. Natl. Bur. Stand. **59**, 249–254 (1957); see also, M. Greenspan and C. E. Tschiegg, ''Tables of the speed of sound in water,'' J. Acoust. Soc. Am. **31**, 75–76 (1959).
[6] W. D. Wilson, ''Speed of sound in distilled water as a function of temperature and pressure,'' J. Acoust. Soc. Am. **31**, 1067–1072 (1959).
[7] A. J. Barlow and E. Yazgan, ''Phase change method for the measurement of ultrasonic wave velocity as a determination of the speed of sound in water,'' Br. J. Appl. Phys. **17**, 807–819 (1966).
[8] J. Kestin and J. V. Sengers, ''New international formulation for the thermodynamic properties of light and heavy water,'' J. Phys. Chem. Ref. Data **15**, 305–320 (1986).
[9] H. Preston-Thomas, ''The international temperature scale of 1990 (ITS-90),'' Metrologia **27**, 3–10 (1990).
[10] N. Bilaniuk and G. S. K. Wong, ''Speed of sound in pure water as a function of temperature,'' J. Acoust. Soc. Am. **93**, 1609–1612 (1993) and Erratum therein **99**, 3257 (1996).
[11] S. Ernst, W. Marczak, and R. Manikowski, ''Temperature fluctuations and gradients in a liquid sample as a source of errors in ultrasound velocity measurements,'' Acoust. Lett. **16**, 177–183 (1993).
[12] J. Czermiński, A. Iwasiewicz, Z. Paszek, and A. Sikorski, *Statistical Methods in Applied Chemistry* (PWN, Warszawa; Elsevier, Amsterdam, 1990), 1st ed., Chap. 6, p. 351.
[13] G. S. Kell, ''Effects of isotopic composition, temperature, pressure, and dissolved gases on the density of liquid water,'' J. Phys. Chem. Ref. Data **6**, 1109–1131 (1977).
[14] W. N. Christiansen, R. W. Crabtree, and T. H. Laby, ''Density of light water: Ratio of deuterium to hydrogen in rain water,'' Nature (London) **135**, 870 (1935); see also Ref. 15.
[15] G. S. Kell, ''Thermodynamic and transport properties of fluid water,'' in *Water. A Comprehensive Treatise*, edited by F. Franks (Plenum, New York, 1972), 1st ed., 4th printing (1983), Vol. 1, Chap. 10, pp. 376–378.

The influence of low-frequency instrumentation response on rocket noise metrics

Sally A. McInerny
Department of Mechanical Engineering, The University of Alabama, Tuscaloosa, Alabama 35484-0276

Jan K. Francine, Brent S. Stewart, and Philip H. Thorson
Hubbs-Sea World Research Institute, 2595 Ingraham Street, San Diego, California 92109

(Received 28 February 1996; revised 19 June 1997; accepted 11 August 1997)

The influence of the low-frequency roll-off and phase distortion of sound level meters (SLMs) on rocket noise metrics based on measured data is studied. Specifically, during the take-off of a space launch vehicle (SLV) called the Taurus, environmental sound pressure was measured with a carrier microphone system. This system has a virtually flat frequency response at low frequencies. In this study, the low-frequency response of SLMs was simulated using high pass filters. Two sets of filters with the same response magnitude were used: with and without phase distortion. It was shown that, for SLVs with small engines (e.g., the Delta, Scout, and Taurus), the low-frequency distortion resulting from measurements with (simulated) SLMs is apparent only in sound pressure power spectral densities (PSDs) of high resolution at frequencies below 10 Hz. This is where PSD levels are already more than 10 dB below the peak levels. There is no significant effect on the skewness of the time domain data nor on the time-energy averaged sound pressure level (SPL). It is shown that for SLVs with large engines the low-frequency response of the typical SLM results in a significant increase in the skewness of the data as peak (positive) shock pressures are increased. Comparisons of the effects of the two high pass filters indicate that the phase of the frequency response is a critical factor in the distortion of the data. [S0001-4966(97)06211-5]

PACS numbers: 43.58.Fm, 43.50.Yw, 43.66.Yw, 43.28.Dm [SLE]

INTRODUCTION

As part of an effort to monitor the effect of the sounds produced by rocket launches on the demographics of endangered species that breed and nest in the vicinity of launch sites, the United States Air Force commissioned the measurement of noise levels during launches of several different vehicles. One federal wildlife regulatory agency requested that the measurements be accurate down to 0.5 Hz, because some avian species use infrasound to monitor for predators. Accurate A-weighted SPLs were also requested, as these were considered important in models of animal hearing sensitivity and susceptibility to noise-induced trauma. Noise was recorded during the launches of a Peacekeeper missile, and Delta, Scout, and Titan IV space launch vehicles (SLVs).

Instrumentation used for that series of launches included Bruel and Kjaer (B&K) type 4155 microphones attached to B&K type 2230 precision sound level meters (SLMs) with the ac output used as an input to TEAC digital audio tape (DAT) recorders. Analyses of the recorded data were provided in reports to wildlife regulatory agencies (Ref. 1) and published in Refs. 2 and 3. This extensive database is used to estimate noise levels at different locations during launches of these and other SLVs of similar thrust.

The high sound levels produced by rockets results in nonlinear propagation. Reference 4 proposes to compare trends predicted by nonlinear acoustic propagation codes with those seen in the extant recorded data. That data, which was taken with commercially available SLMs, may not reflect accurately low-frequency components because of the frequency response of these instruments. Since the low-frequency components are critical to the accurate representation of the time domain waveforms, the value of such comparisons could be questionable.

The launch data measured with the SLMs may be part of a valuable database for both launch noise predictions and the investigation of nonlinear acoustic propagation effects, but only if the influence of the instrumentation system on the recorded data is either insignificant or clearly understood. To this end, we engaged in a study of the influence of instrument low-frequency roll-off and phase distortion on rocket noise metrics. The baseline data set was recorded during the launch of a Taurus SLV (which uses a modified Peacekeeper engine) using a carrier microphone system. We evaluated the data using several measures, three of which are discussed here: (1) time-energy averaged SPL (often written as L_{eq}); (2) crest factor (peak to rms ratio); and (3) skewness (the third moment of the time domain data).

The time-energy averaged sound level is used to compare average noise levels. A higher crest factor indicates larger deviation from the mean-square level. Skewness is a statistical measure of the distribution of sound pressure. Nonzero skewness indicates that the noise amplitudes do not follow a Gaussian distribution. A positive skewness indicates a greater number of high-level positive pressure peaks than high-level negative pressure peaks. All of the rocket noise data measured using the standard SLMs were positively skewed, with skewness values that ranged from 0.02 to 0.55.[3] A positive skewness is associated with the presence of

shocks, perceived by the human ear as ''crackle,'' in supersonic jet noise.[5]

Just as the mean-square value of a zero-mean signal (i.e., its variance) is related to the one-dimensional integral of its power spectrum, the skewness of a signal is related to the two-dimensional integral of its bi-spectrum.[6] The bi-spectrum, as well as other higher order spectra, are used for the characterization of nonlinear and non-Gaussian signal properties.[7,8] Applications of bi-spectral analysis to high intensity acoustic noise can be found in Refs. 9 and 10. The effect of low-frequency instrumentation response on the higher order spectra of the recorded data are not examined here and remain to be investigated. In this paper, we examine the effect of low-frequency roll-off on the skewness of the recorded data.

I. DATA SET AND ANALYSIS

The B&K Type 2230 SLM, type 4155 microphone and DAT recorder that together were used to record launch noise data from the Peacekeeper, Delta, Scout, and Titan IV, will hereafter be referred to as the ''SLMs.'' The frequency response of a commercial precision sound level meter, like the type 2230, is determined by a combination of microphone, preamplifier, and meter responses. Usually the manufacturer does not document the frequency response below 10 Hz. We wanted to compare the measured data from a SLM to that measured with a more accurate instrument at low frequencies.

The B&K type 2631 microphone carrier system with B&K type 4147 microphone and DAT recorder, that were used for the Taurus SLV launch noise measurement, will be referred to as the ''carrier system.'' [11] The type 2631 carrier system, no longer manufactured, uses a 10-MHz carrier frequency instead of a constant voltage to polarize the microphone. This design permits accurate measurements down to 0.01 Hz (± 1 dB) when used with the type 4147 microphone. The TEAC DAT recorder, used to record the noise data, has a flat frequency response to dc.

As is the case with aircraft take-off noise, the amplitude of the noise generated by a SLV varies substantially over a short period of time. This noise is frequently characterized by metrics averaged over the time period when the overall sound pressure level (OASPL) is within 6 dB or 10 dB of its maximum. We refer to these time periods as the 6 dB down and 10 dB down time periods (see Fig. 1). Launch noise measured during these periods of peak acoustic levels is rich in shocks, as shown by the abrupt increases in pressure in the waveforms of Fig. 2.

II. SIMULATED INSTRUMENTATION RESPONSES

We followed the approach used by Kerry, Ford, and James in a study of blast noise.[12] We used the essentially undistorted data measured with the carrier system and simulated the SLM responses with high pass filters. Filters with the same response magnitude, but with and without phase distortion, were utilized. We analyzed the recorded data and present here a comparison of the metrics calculated from the

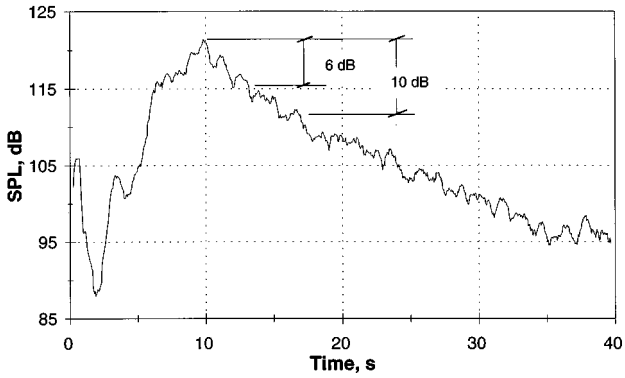


FIG. 1. Sound pressure levels (in dB *re:* 20×10^{-6} Pa) at 2.2 km from Taurus pad. Levels are 0.5-s linear rms averages calculated from the digital data.

filtered data with those of the unfiltered data and then draw inferences regarding the recordings of launch noise made with the SLM systems.

Prior to any further processing, we calculated the following metrics for the Taurus data (measured with the carrier system) over the 6 dB down period: the time-energy averaged SPL, the skewness coefficient, and the crest factor. The crest factor is defined as

$$\text{crest factor} = 20\log_{10}\frac{\max p_i}{p_{\text{rms}}} \quad \text{in dB}.$$

The skewness coefficient is defined as

$$\text{skewness coefficient} = \sum_{i=1}^{n} \frac{(p_i-\mu)^3}{\sigma^3},$$

where p_i is the instantaneous pressure amplitude, μ is the mean sound pressure (nominally zero), and σ is the standard deviation (here, $\sigma = p_{\text{rms}}$). The summation $i=1,n$ encompasses the number of instantaneous data values in the 6 dB down averaging period.

The Taurus raw time data were subsequently digitally low pass filtered at 1 kHz, using a 120-coefficient, flat-phase finite impulse response filter with a Gibbs factor of 1.0. In order to minimize calculation times and avoid numerical problems that were previously encountered when the spectral bandwidth (BW) was very small relative to the Nyquist fre-

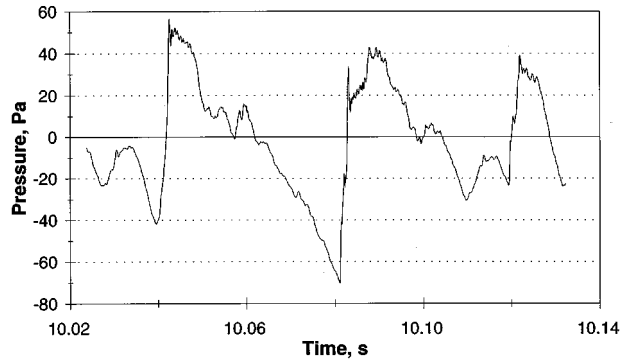


FIG. 2. Example of shock waveforms in the Taurus noise data.

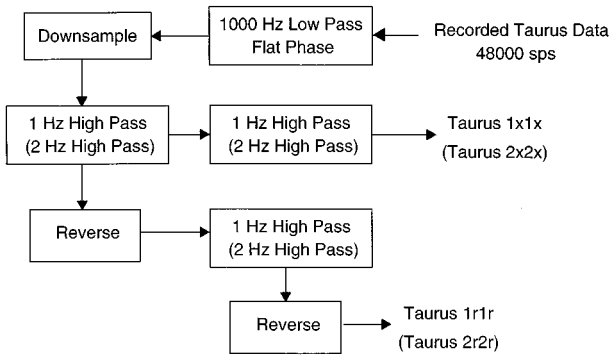


FIG. 3. Designation of filtered data sets. High pass filters were 1 pole Butterworth filters with the indicated cutoff frequencies.

quency, the data were then down sampled by a factor of 10. (After downsampling, the sampling rate of the data was 4800 samples per second.)

To simulate the influence of SLMs on the Taurus data, the low passed, down sampled Taurus data were then processed as shown schematically in Fig. 3. The procedure in which the data are high pass filtered and reversed twice results in a filter that has no phase distortion. The magnitude of the composite filter response is the same as that of the twice-filtered data without the reversals. When a 1-pole Butterworth filter with a 2-Hz cutoff frequency is used, the procedure results in a composite $2x2x$ filter (see Fig. 3 for an explanation of the filter designations) that has a frequency response magnitude that closely approximates those measured for the SLMs exclusive of the microphones[1] (see Fig. 4). Microphone manufacturers do not commonly provide any low frequency (below 20 Hz) information on amplitude or phase response. Therefore, we are not able to assess the additional influence of the microphone. Note that (from Fig. 3) the $1r1r$ and $2r2r$ filters have the same response magnitudes as the $1x1x$ and $2x2x$ filters, respectively, but have no phase distortion.

Of the four rockets measured with the SLMs, the Titan IV has by far the largest engines, hence generates noise of significantly lower frequency. The engine nozzle exit diameters of the Titan IV and Taurus are 3.35 m and 1.53 m, respectively. Both launch vehicles have solid rocket engines with approximately the same exit velocities. To simulate the noise produced by the larger Titan IV engines, the Taurus data were time scaled by a factor of 3.35/1.53, effectively shifting frequencies down by a factor of 0.45. This simulated Titan IV data set was then high pass filtered as described in the previous paragraph.

Skewness coefficients and rms pressures were calculated for all of the filtered and unfiltered data sets over short duration, sliding time frames. For the Taurus data sets, a 0.5-s time averaging period and 90% averaging overlap was used. For the simulated Titan IV data (i.e., the time scaled Taurus data), to maintain equivalent statistical degrees of freedom (an equal number of digital data points in each average) a 1.1-s averaging time was used. Averages over the 6 dB and 10 dB down time periods were calculated from the short duration averages.

For the narrow-band sound pressure PSDs, we chose a 0.5-Hz BW and used Hanning windowing for all of the filtered and unfiltered data sets. Because a 0.5-Hz BW FFT is calculated from a relatively long 2.0-s time sample, we used the 10 dB down averaging period to increase the number of FFTs in the average. (A 90% overlap was used to minimize the unused segment of data at the end of the 10 dB down time period.)

III. RESULTS

As seen in Table I, the time-energy averaged SPL, skewness coefficient, and crest factor, calculated for the Taurus

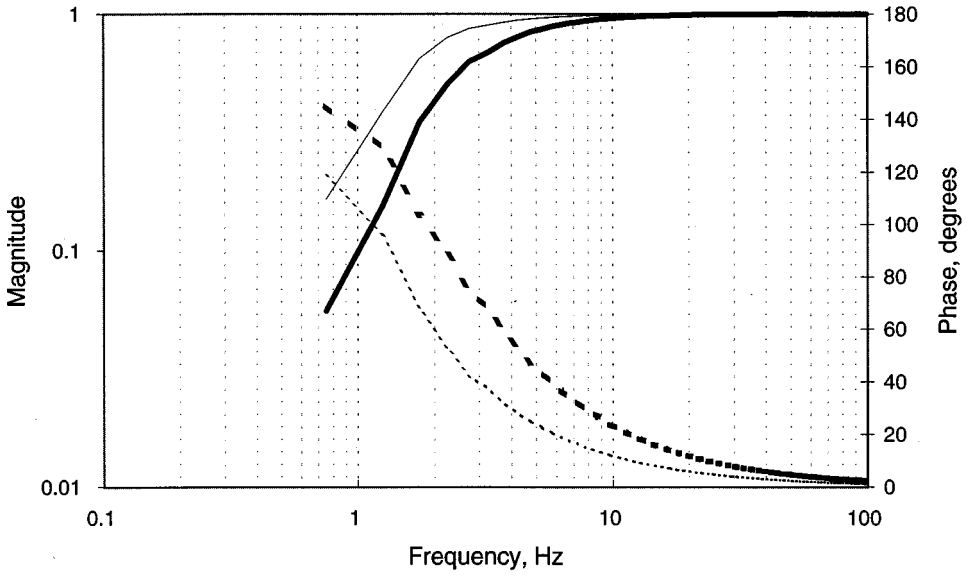


FIG. 4. Frequency response of composite filters represented in Fig. 3. Solid lines are used for the magnitude; dashed lines for the phase. Light lines represent the $1x1x$ filter, heavy lines represent the $2x2x$ filter.

TABLE I. Rocket noise metrics averaged over the 6 dB down period. Delta, Scout, Peacekeeper, and Titan IV data were measured with SLM system; Taurus data were measured with carrier system.

Vehicle	Distance, d (km), from launch pad	d/De[a]	L_{eq} (dB)	Crest factor (dB)	Skewness coefficient
Delta	0.46	400	128.7	13.0	0.35
	0.61	530	128.3	12.4	0.071
	0.92	800	122.8	14.8	0.041
Scout	0.36	420	125.7	13.2	0.12
	0.64	740	118.9	14.8	0.14
	0.94	1100	115.2	15.8	0.021
	1.22	1410	111.5	14.7	0.063
Peacekeeper	0.31	210	134.9[b]	>12.9[b]	0.21
	0.61	420	128.2	12.8	0.13
	1.22	840	123.3	13.9	0.43
Titan IV	0.82	245	139.5	14.4	0.55
	2.00	605	129.0	11.3	0.42
	3.41	1020	127.0	13.0	0.38
	5.79	1725	119.0	12.6	0.074
	13.15	3920	109.2	15.5	0.32
Taurus	2.24	1470	117.9	13.2	0.18

[a]d/De is the distance from the pad normalized by the engine(s) exit diameter.
[b]Five positive pressure peaks were clipped at the input to the DAT recorder.

data (measured with the carrier system) over the 6 dB down period, are comparable to those previously obtained for the Peacekeeper, Delta, Scout, and Titan IV data (measured using SLMs). When compared in this fashion, there does not appear to be anything that clearly points to the Taurus data as being different or nongeneric.

The results shown in Table II provide a more controlled comparison of the effects of low-frequency instrumentation response on recorded rocket noise. Table II shows that the low-frequency roll-off of a typical SLM (as approximated by the $2x2x$ filter) has a very minor influence on measured time-averaged SPLs over the 6 dB down period. This is true

TABLE II. Comparison of metrics for data sets with and without high pass filtering. All data were low-pass filtered and downsampled before further processing. See Fig. 3.

Data set	Minimum pressure[a] (Pa)	Maximum pressure[a] (Pa)	6 dB down skewness	6 dB down L_{eq} (Pa)	10 dB down skewness	10 dB down L_{eq} (Pa)
Taurus	−69.84	65.61	0.172	117.9	0.092	116.5
$1x1x$	−65.93	66.25	0.193	117.9	0.112	116.4
$2x2x$	−61.05	66.31	0.215	117.9	0.133	116.4
$1r1r$	−69.61	66.35	0.174	117.9	0.094	116.4
$2r2r$	−69.14	64.72	0.178	117.9	0.097	116.4
Simulated Titan IV	−69.84	65.61	0.175	118.0	0.093	116.5
$1x1x$	−60.08	66.28	0.222	117.9	0.138	116.4
$2x2x$	−55.62	67.03	0.267	117.8	0.187	116.2
$1r1r$	−69.03	64.58	0.180	117.9	0.098	116.4
$2r2r$	−67.48	62.67	0.186	117.8	0.102	116.2

[a]Instantaneous digital sound pressures. Maximums were 0.7 dB higher before low pass filtering.

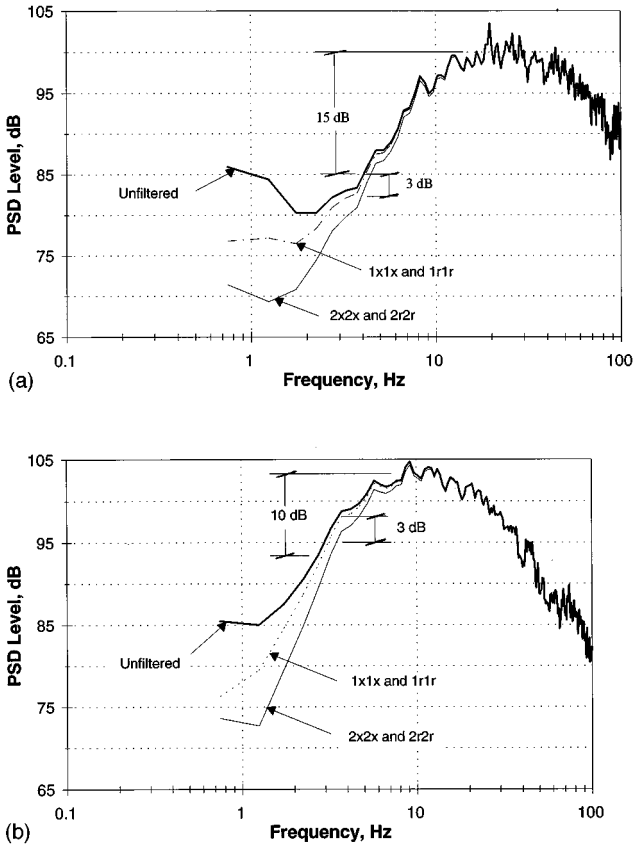


FIG. 5. 0.5-Hz resolution sound pressure spectral densities of filtered and unfiltered: (a) Taurus; and (b) simulated Titan IV (time scaled Taurus) data. Heavy solid line is the unfiltered data; dashed line is the $1x1x$ filtered data; the light solid line is $2x2x$ filtered data. Spectral levels are averages over the 10 dB down time period, see text. Units are dB *re:* 20×10^{-6} Pa/Hz.

even for data with a frequency distribution typical of the Titan IV.

The influence of the applied filters on the 6 dB down Taurus and Titan IV (time scaled Taurus) spectra is shown in Fig. 5(a) and (b). The Taurus sound pressure PSD levels are down 15 dB from the peak before the $2x2x$ filter significantly affects the spectrum. In the case of the simulated Titan IV data, the $2x2x$ filter results in a 3-dB decrease in PSD levels within 10 dB of the maximum. [The peak PSD levels for the simulated Titan IV data in Fig. 5(b) are higher than those of the Taurus in Fig. 5(a), because time scaling reduced the effective BW while the energy was preserved.]

In Table II, the skewness of the data can be seen to increase with an increase in the cutoff frequency of the high pass filter. The phase distortion of the system is the dominant factor rather than the magnitude of the frequency response. To emphasize the importance of low-frequency phase distortion, plots of short duration skewness coefficients for the simulated Titan IV data sets are compared in Fig. 6. There are only small differences between the results for the $2r2r$ (no phase distortion) filtered data and the unfiltered data. The differences between the $2x2x$ (with phase distortion) data and the unfiltered data are much more dramatic. (Recall that the $2r2r$ and $2x2x$ filters have identical frequency response magnitudes.)

The influence of the simulated low-frequency instru-

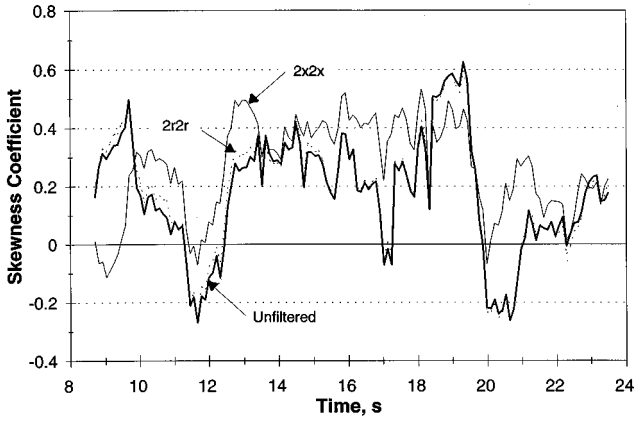


FIG. 6. Short duration skewness coefficients of the filtered and unfiltered simulated Titan IV data. Heavy solid line represents the unfiltered data; dashed line the $2r2r$ filtered data; and the light solid line the $2x2x$ filtered data.

mentation roll-off (with phase distortion) on maximum and minimum instantaneous sound pressures is consistent with the increase in skewness. Both positive pressure maxima and negative pressure minima are positively increased. In Fig. 7, a segment of the simulated Titan IV time domain data (i.e., the time scaled Taurus data) are plotted. The unfiltered data are shown as a heavy solid line, while the $2x2x$ filtered data (with phase distortion) are shown as a light solid line. It can be seen that the phase distorting filter results in increased peak positive pressures at the top of the shocks and shallower negative pressure troughs ahead of the shocks.

IV. CONCLUSION

Noise data, measured with a B&K type 2631 carrier system and a B&K type 4147 microphone, during the launch of a Taurus SLV were used as a baseline in a study of the effect of low-frequency instrumentation response on rocket noise metrics. Of particular interest is the effect of the low-frequency roll-off and phase distortion of SLMs that were previously used to measure noise during the launches of Delta, Peacekeeper, Scout, and Titan IV rockets. To study

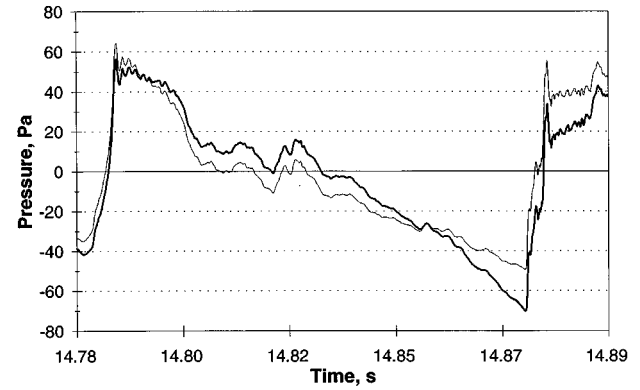


FIG. 7. Waveform in the filtered and unfiltered simulated Titan IV data. The heavy solid line represents the unfiltered data; the light solid line represents the $2x2x$ filtered data.

and simulate the low-frequency instrumentation response of these SLMs (exclusive of microphones) we applied high pass filters to the baseline data.

The effect of the SLM (as simulated by high pass filters) on the Taurus data was apparent only in high resolution sound pressure spectral densities (1.0-Hz and 0.5-Hz analysis BWs) at frequencies below 10 Hz. At these frequencies, spectral levels were down more than 10 dB from the peak levels in the spectrum. No significant effect was found on skewness or on the time averaged SPLs over the 6 dB down time period. These results are expected to hold for the noise of other launch vehicles with small engines, such as the Peacekeeper, Delta, and Scout rockets.

However, the effect of the SLM (as simulated by high pass filters) on the noise generated by larger engines, as represented by the simulated Titan IV data (time scaled Taurus data), was a significant increase in the skewness of the data. Also, peak positive shock pressures were increased. The low-frequency phase response of the measurement system was found to be the dominant factor in the distortion of the data. Although the influence of the low-frequency instrumentation roll-off was very evident in 0.5-Hz BW spectra, the influence on the time-energy averaged SPL over the 6 dB and 10 dB down periods was insignificant.

It is clear that measurements of rocket launch noise made with standard SLMs understate the very low-frequency spectral levels. But we found little effect otherwise on the characteristics of the recorded noise in the case of the Delta, Peacekeeper, and Scout rockets. However, for the Titan IV, recorded levels within 10 dB of the peak level in the spectrum are affected by the roll-off in the low-frequency instrumentation response. Skewness coefficients published for the Titan IV data (see Table I) are likely to be overstated and the time-energy averaged SPLs somewhat understated.

Depending on the distance of exposed wildlife to rocket launch pads and the size of the SLV engines, these measurement differences could have substantial effects on predictions and assessments of the effects of the noise on animal auditory function.

[1] Reports (a)–(c) listed below were authorized for public release: (a) S. A. McInerny, ''Peacekeeper Launch Sound Levels at 1000, 2000, and 4000 Feet from the Pad,'' Aerospace Report ATR-93 (3566)-1, Dec. 1992; (b) S. A. McInerny, ''Delta Launch Sound Levels at 1500, 2000, and 3000 Feet from the Pad,'' Aerospace Report ATR-93 (3566)-2, Jan. 1993; (c) S. A. McInerny, ''Scout Launch Sound Levels at 1000, 2000, 3000 and 4000 Feet from the Pad,'' Aerospace Report No. ATR-93 (3566)-3, Mar. 1993; (d) T. T. Do, ''Titan IV Launch Sound Levels at 2700, 6800, 11200, 16800 and 43129 Feet From the Pad,'' Aerospace Tech. Rep. TOR-94 (4566)-1, Jan. 1994.

[2] S. A. McInerny, ''Launch Vehicles, Part I-Overall Levels and Spectral Characteristics,'' J. Aircr. **33**, 511–517 (1996).

[3] S. A. McInerny, ''Launch Vehicles, Part II—Statistics of the Time Domain Data,'' J. Aircr. **33**, 518–523 (1996).

[4] S. A. McInerny, J. K. Wickiser, and R. H. Mellen, ''Rocket noise propagation,'' to be presented in the Symposium on Radiation and Scattering of Sound, IMECE, 16–21 November 1997, Dallas, TX.

[5] J. E. Ffowcs Williams, J. Simson, and V. J. Virchis, ''Crackle: An annoying component of jet noise,'' J. Fluid Mech. **71**, 251–271 (1975).

[6] C. L. Nikias and A. P. Petropulu, *Higher Order Spectra Analysis, A Nonlinear Signal Processing Framework*, Prentice Hall Signal Processing Series (Prentice-Hall, Englewood, NJ, 1993).

[7] C. L. Nikias and J. M. Mendel, ''Signal processing with higher-order spectra,'' IEEE Trans. Signal Process. **20**, 10–37 (1993).

[8] M. J. Hinich, ''Testing for Gaussianity and linearity of a stationary time series,'' J. Time Series Anal. **3**, 169–176 (1982).
[9] R. H. Mellen and I. A. Leykin, ''Nonlinear dynamics of wind waves: MultiFractal phase/time effects,'' Nonlinear Process. Geophys. **1**, 51–56 (1994).
[10] S. N. Gurbatov, A. N. Malakhov, and Pronchatov-Rubtsov, ''Application of higher-order spectra in problems of diagnosis of strong acoustic noise,'' Sov. Phys. Acoust. **33**, (1987).
[11] B. S. Stewart, J. K. Francine, and P. H. Thorson, ''Taurus launch at Vandenburg Air Force Base, 13 March 1994; Sound levels and behavioral responses of Harbor Seals (Phoca vitulina richardsi) at Purisma Point and Rocky Point,'' Hubbs-Sea World Research Institute Technical Report No. 94-252, submitted to USAF SMC/CEW, 20 May 1994.
[12] G. Kerry, R. D. Ford, and D. James, ''Bandwidth limitation effects on low-frequency impulse noise prediction and assessment,'' Symposium on Aircraft Noise Abatement, sponsored by the NATO Committee on the Challenges of Modern Society, 16–20 May 1994, Baltimore, MD.

Automated evaluation of acoustic talker direction finder algorithms in the varechoic chamber

Dennis R. Morgan
Bell Laboratories, Lucent Technologies, 700 Mountain Avenue, Murray Hill, New Jersey 07974-0636

Vipul N. Parikh[a)]
Voxware, Inc., 305 College Road East, Princeton, New Jersey 08540

Cecil H. Coker
Bell Laboratories, Lucent Technologies, 700 Mountain Avenue, Murray Hill, New Jersey 07974-0636

(Received 21 January 1997; revised 28 July 1997; accepted 1 August 1997)

Acoustic talker direction finders have potential applications to camera pointing for teleconferencing and to microphone array beam steering for audio communication and voice processing systems. This paper describes a laboratory setup and computer interface that was developed for testing talker direction finder algorithms in the Bell Labs Varechoic chamber, a room with computer-controlled absorbing panels. The procedure exploits the full capability of the facility by automatically stepping through a sequence of room panel configurations, outputting a digital speech signal, running the processor, and collecting the data. The advantage of this technique is that it allows for testing under a multitude of different acoustic conditions in the same physical location, thereby enabling a general characterization of the algorithm under evaluation. As an example of the technique, we have implemented the Fischell–Coker talker direction finder algorithm using real-time C-code running on an SGI workstation, which is the same machine that is used to orchestrate the automatic testing procedure. [S0001-4966(97)03911-8]

PACS numbers: 43.58.Ta, 43.66.Qp, 43.60.Gk, 43.60.Qv [SLE]

INTRODUCTION

This paper is concerned with locating a talker in a reverberant room using measurements of the acoustic field. Applications of such a technique include video camera pointing and microphone array beam steering. Camera pointing would be useful in multimedia video telecommunication and security monitoring systems, where it is necessary to locate and track an active talker. Steering a high-gain microphone array toward an active talker would be useful for reducing noise and reverberation, thereby enhancing speech quality and enabling reliable automatic speech recognition without using a close microphone.

The basic concept of an acoustic talker direction finder is to process the speech signals obtained from a collection of two or more microphones. This problem is a difficult one for several reasons. First of all, speech itself is highly variable from talker to talker, and is furthermore nonstationary even for a specific talker. Also, there are room acoustic degradations, including multipath/reverberation and variability of response from room to room due to geometry, absorption, temperature, and the like. All of this makes it difficult to invent a universal technique that will work reliably over a wide range of conditions.

There have been several approaches to finding the location of a talker. The simplest method is to cross-correlate signals from two or more microphones and select the peak location as an estimate of time delay, which determines the angle of arrival for a plane wave signal. Two such microphone pairs then enable position estimation by triangulation. The basic technique can be enhanced by hyperbolic curve fitting of multiple delay estimates[1] or by linear 3-D intersection of multiple bearing lines.[2,3] These techniques work well in free space. However, a fundamental weakness of all such cross-correlation based algorithms is an inability to cope well with the effects of multipath and reverberation in an enclosed room.[4] Some improvement may be gained by cepstral prefiltering,[5] however, shortcomings still remain.

More sophisticated techniques[6,7] are based on complex models of the human binaural hearing mechanism, and embellish the cross correlation by processing in multiple bands with contralateral inhibition,[8,9] which is a mechanism introduced for explaining the so-called ''law of first wavefront'' or ''precedence effect.'' [10] However, there are still problems in a reverberant environment and further research on the precedence effect has been suggested.[6] Even more complicated models can be obtained by combining time delay with amplitude and onset cues[11] and could form the basis for a robust location algorithm.

A simple talker direction finder algorithm was suggested by Coker and implemented by Fischell and Coker[12] that is based on a heuristic imitation of the precedence effect. The basic idea is to identify the onset of pitch periods and to cross correlate these within a certain acceptance window.

No matter what type of algorithm is used, a need exists for efficient and comprehensive performance evaluation over a range of representative acoustic conditions, and that is the main focus of this paper. The recently constructed Bell Labs Varechoic chamber was used as a basis for this work. This facility consists of a 6.71 m×5.94 m×2.74 m (264 in.

[a)]Bell Laboratories summer employee 1995.

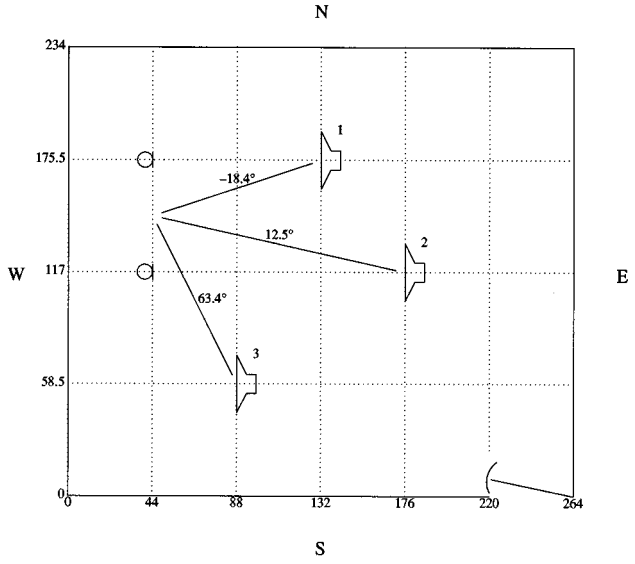


FIG. 1. Varechoic chamber floor plan (coordinate values measured in inches); loudspeaker sources at (132, 175.5), (176, 117), and (88, 58.5); TDF centered at (44, 146.25); microphones at (44, 117) and (44, 175.5).

$\times$234 in.$\times$108 in.) room with electronically controlled panels that vary the acoustic absorption of the walls, floor, and ceiling.[13] An automated computer-controlled process is described that exploits the full capability of the facility by stepping through a sequence of room panel configurations, outputting a digital speech signal, running an arbitrary talker direction finder, and collecting the data. In this way, tedious and error prone measurement campaigns are replaced by complete automation, thereby greatly relieving the drudgery and improving quality control. This is particularly useful when the experiment must be repeated because of some defect in the apparatus or physical setup, or for trial of a new experiment where comparison will be made with past measurements.

As an example of the process, an implementation of the Fischell–Coker talker direction finder (TDF) algorithm[12] was developed using real-time C-code running on an SGI workstation. The automated testing process was also implemented on the same SGI machine, which was interfaced with the Varechoic chamber controller PC through a networked disk file.

In Sec. I, the room setup in the Varechoic chamber is described. Section II then details the experimental procedure and the results are reported in Sec. III. The Fischell–Coker C-code implementation developed for these experiments is sketched in the Appendix.

TABLE I. Average 60-dB reverberation time in the 400–1600 Hz band as a function of the percentage of open panels.

Percentage open panels	Reverberation time (ms)
0	740
25	375
50	250
100	150

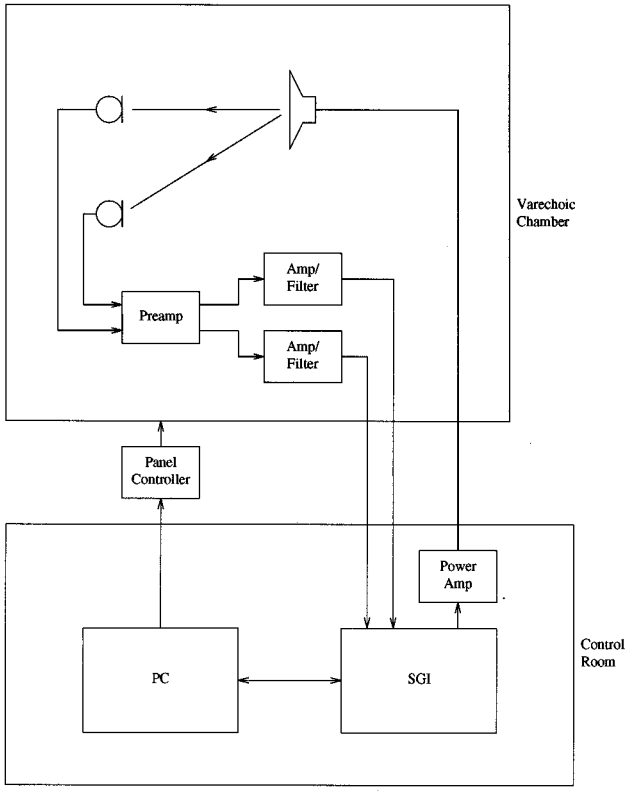


FIG. 2. Overall setup of TDF experiment.

I. ROOM SETUP

Figure 1 shows a diagram of the floor plan layout, which is composed of 96 individually controllable panels, organized into 12 bays of 8 panels each (the dotted lines in the figure represent visual seams that subdivide each bay into two N-S sections). Similar arrangements organize 11 bays of 88 ceiling panels, 6 bays of 48 panels for each of the North and South walls, and $5\frac{1}{2}$ bays of 44 panels for each of the East and West walls, resulting in a grand total of 368 panels. Each panel (except for one on the South wall used for ventilation) consists of two perforated sheets whose holes, if aligned, expose sound absorbing material behind, but if shifted to misalign, form a highly reflective surface. The panels are individually controlled so that the holes are either fully open (absorbing state) or fully closed (reflective state). Therefore, by varying the binary state of each panel in any combination, a plethora (2^{368}) of different room characteristics can be simulated, more of which will be said later in Sec. II.

For convenience, positions in the floor plan of Fig. 1 will be designated by (x,y) coordinates corresponding to inches along the (South, West) walls, corresponding to original plan specifications. The regimen of tests reported in this paper were conducted with the TDF receiver centered on the (44, 146.25) coordinate, corresponding to the midpoint between the two microphones in Fig. 1. Talker sources were simulated by placing a Celestion SR loudspeaker on a stand located in one of the three positions shown in Fig. 1. These positions will be referred to as ''room1'' (132, 175.5), ''room2'' (176, 117), and ''room3'' (88, 58.5). The positions

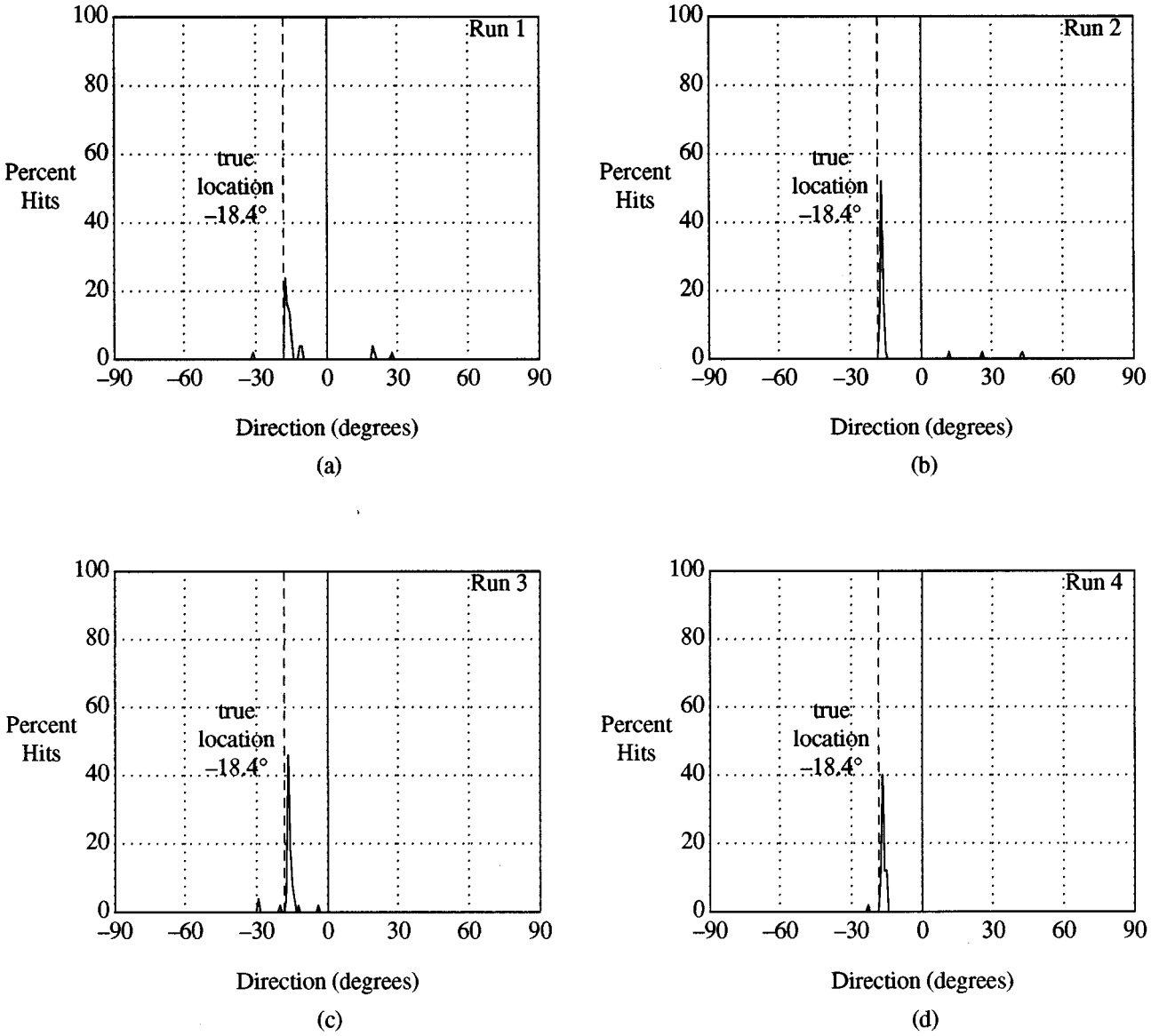


FIG. 3. Run-to-run variability; room1 talker position, 0% open panels (maximum reverberation). (a) Run 1. (b) Run 2. (c) Run 3. (d) Run 4.

of the TDF and loudspeaker locations were selected for ease in location relative to the panel bay seams and also for obtaining a representative selection of asymmetric geometries.

A total of 10 different panel configurations were selected, comprising 0% and 100% open, and four different random selections each for 25% and 50% open. Reverberation time measurements were made in the 400–1600 Hz band for each of these configurations and the results are summarized in Table I. The reverberation time decreases monotonically as the percentage of open panels increases, because of the increased sound absorption. All of the experiments in this paper are characterized as a function of the percentage of open panels, and with the above table can be related to a desired application roughly described by a given reverberation time.

II. EXPERIMENTAL PROCEDURE

This section describes the procedure used to generate experimental data using a C-code implementation of the Fischell–Coker TDF algorithm,[12] which is described in the Appendix. Figure 2 depicts the overall setup consisting of the Varechoic chamber and control room. A Gateway 2000 486-66V PC and an SGI workstation in the control room were used to change the panel configurations, play a speech file, start the TDF, and collect the data. Once the panel configuration is set by the PC, the SGI plays the speech file through a loudspeaker and digitizes speech signals from two microphones. The speech signals are processed in real time by the SGI, whose output is displayed on the monitor and stored in appropriate files for later analysis. A script file running on the SGI orchestrates all of these events.

A raw speech file was compiled from a speech data base sampled at 16 kHz, and consists of the following three sentences:

''Bobby did a good deed.''
''Do you abide by your bid?''
''A teacher patched it up.''

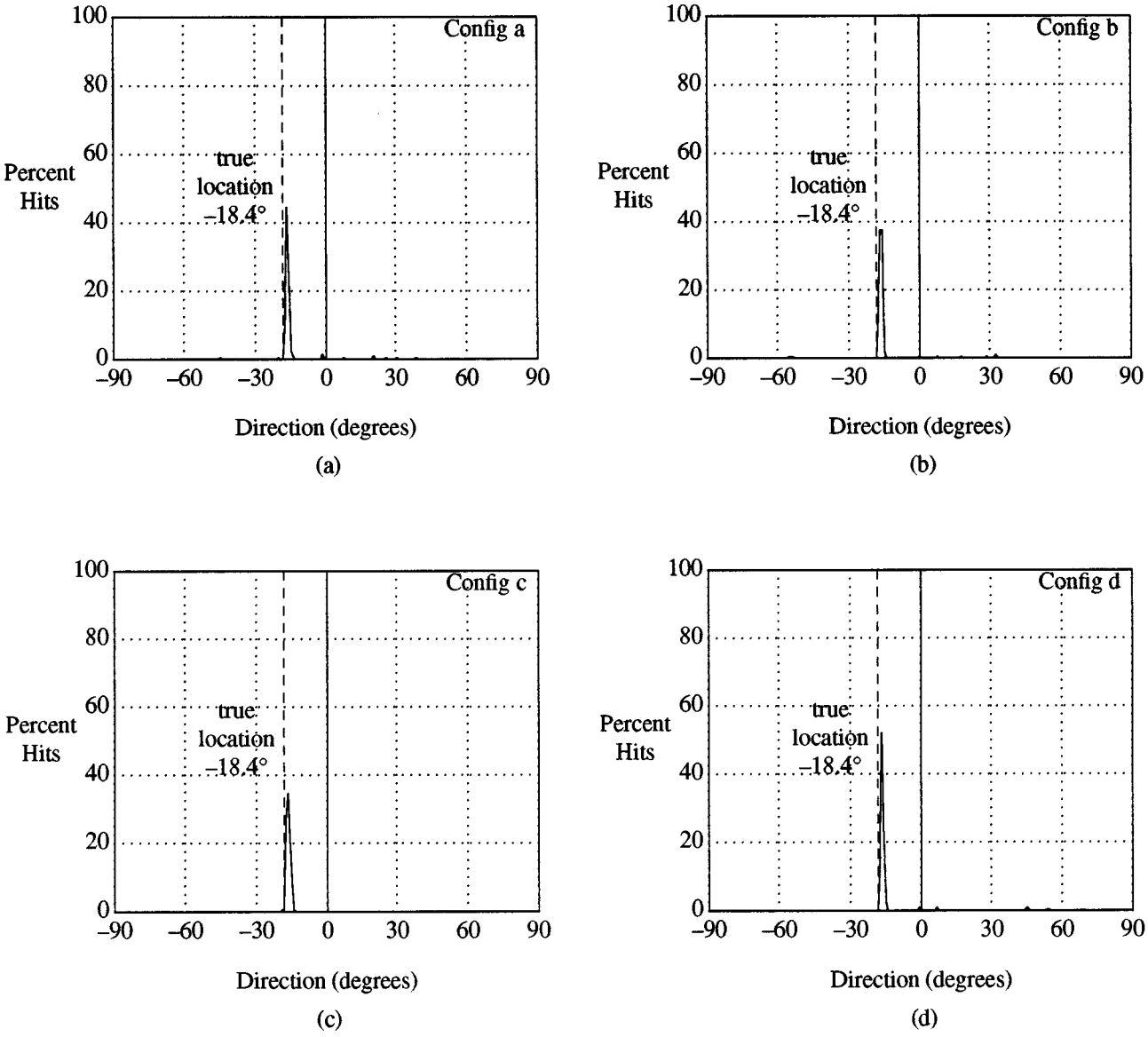


FIG. 4. Panel configuration-to-configuration variability; room1 talker position, 50% open panels, average of 4 runs. (a) Configuration a. (b) Configuration b. (c) Configuration c. (d) Configuration d.

The file comprises 85 278 samples and plays for about 5.3 s at the 16-kHz sampling rate. The speech file was played through the SGI headphone output to a Celestion SR loudspeaker via a Crown amplifier with gain set of −20 dB. Three different positions of the loudspeaker were selected to evaluate the performance of the TDF, as shown in Fig. 1. The positions of the two microphones, also shown in Fig. 1, were kept fixed throughout the experiment. Two B&K type 4011 cardioid microphones were used and were oriented facing to the right, or East, in Fig. 1. The outputs of the microphones were fed to a two-channel AERCO MP2 preamplifier, whose gain was set to 31.5 dB. Each output of the AERCO MP2 goes to a Stanford SR560 amplifier/filter, set to 30-kHz low pass with gain of 5 (14 dB). The Stanford outputs are fed to the stereo line input port of the SGI. This setup provides two-channel input to the TDF implemented on the SGI machine.

For our version of the Fischell–Coker algorithm, the output files are generated one line at a time every 100 ms, and each entry consists of either a location bin number from −80 to 80 or the word ''xx'' if no detection is made. The output files were postprocessed by a routine that counts the total number of lines, forms a histogram of detected bin numbers, converts the bin numbers to angle in degrees, and plots the histograms in terms of the number of hits in each bin as a percentage of the total number of output lines.

III. RESULTS

As a preliminary step, the effect of the algorithm threshold was determined for talker position 1 averaging over eight runs. This process was carried out for 0%, 25%, 50%, and 100% open panel configurations. In each case, the threshold integer (see the Appendix) was varied and receiver operating curves (ROCs) were obtained by plotting the miss probability as a function of the false alarm probability. As a result of

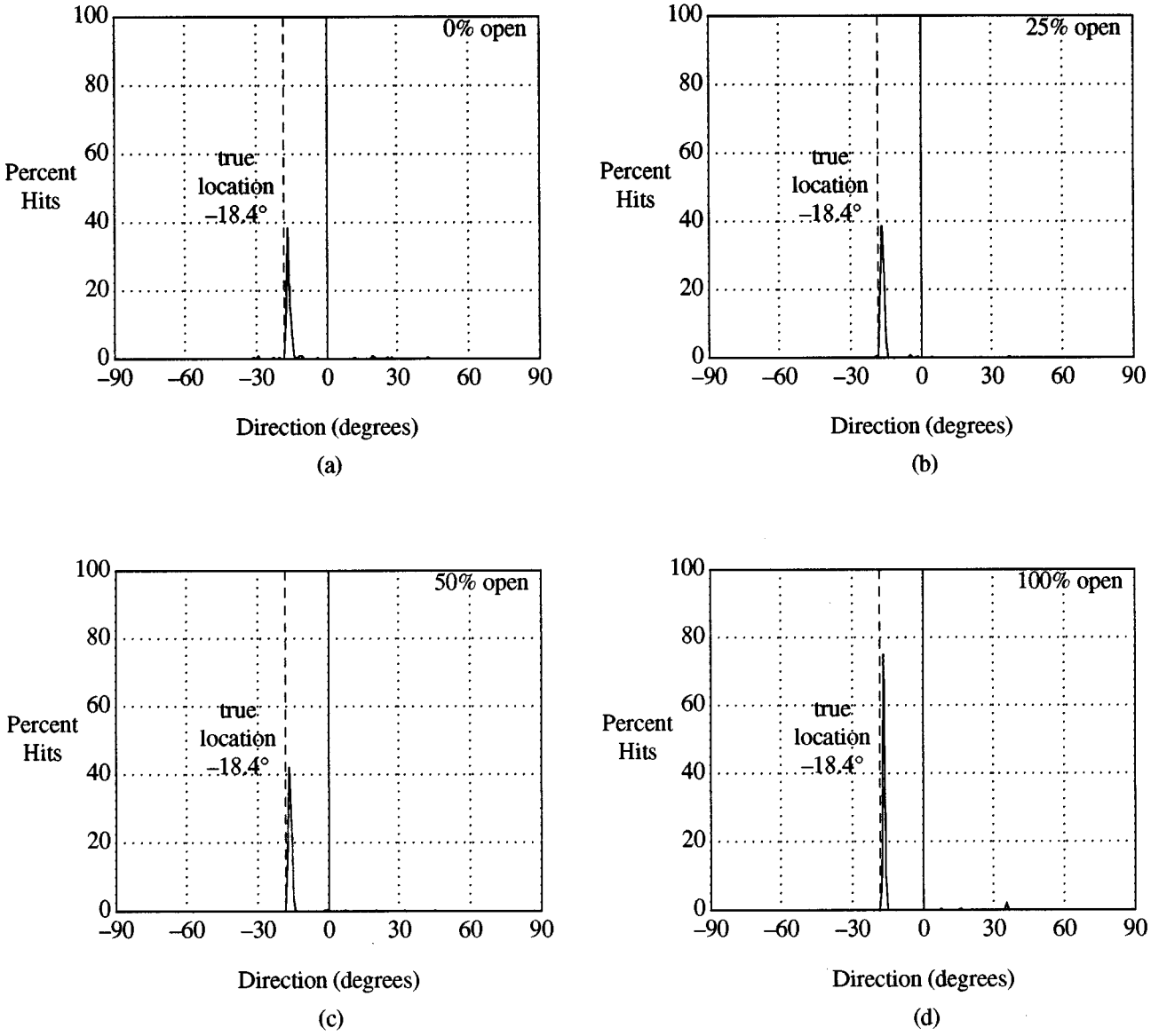


FIG. 5. Panel open-to-open variability; room1 talker position, average over 4 panel configurations (except for 0% open and 100% open), average of 4 runs. (a) 0% open panels (maximum reverberation). (b) 25% open panels. (c) 50% open panels. (d) 100% open panels (minimum reverberation).

this exercise, a single threshold value was selected as the best overall compromise and was used for all of the following experiments.

Figure 3 shows summary histograms for four runs of talker position 1 with 0% open panels (maximum reverberation). As can be seen, there is considerable variation from run to run because of the timing with respect to the four overlapping interval histograms in the algorithm (see the Appendix) and the inherent nonstationarity of the speech signal. For this reason, we decided to average over four runs when plotting summary histograms.

There are four different random panel realizations each for the 25% and 50% open cases. The performance variability over the four configurations is shown in Fig. 4 for the 50% open case with talker position 1 and average over four runs. In the subsequent results, we sometimes choose to average over this variability to form a more general characterization of performance.

Figure 5 shows the performance of the room1 talker position as a function of the percent open panels, averaging over four runs and over four random panel configurations (except for the 0% and 100% open cases). It can be seen that the performance is consistently good when there is any appreciable amount of reverberation (0%–50% open) and excellent when there is little reverberation (100% open).

Finally, the effect of talker location is demonstrated by comparing representative results for one 50% open panel configuration, averaging over four runs. Figure 6 shows the summary histograms for the three talker locations of Fig. 1. Talker location 3 seems to be the most problematic, probably owing to the unfavorable geometry which in this case places reflections off the South wall close to the direct path. The TDF performance was also determined using omnidirectional microphones and appreciable degradation was noted, especially under highly reverberant conditions.

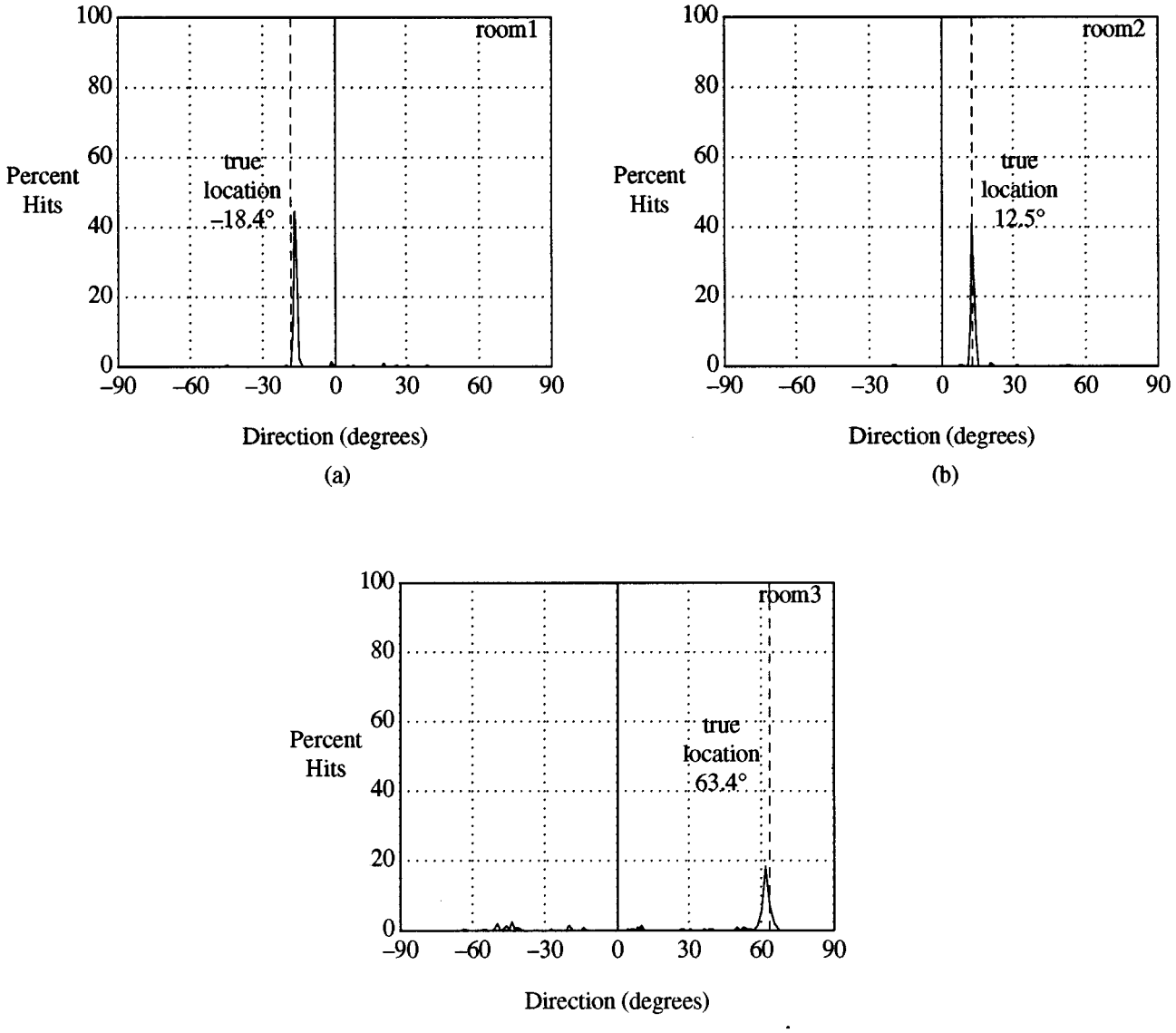


FIG. 6. Talker location-to-location variability; 50% open panels (configuration a), average of 4 runs. (a) Room1 talker position. (b) Room2 talker position. (c) Room3 talker position.

ACKNOWLEDGMENTS

The authors wish to express their thanks to Gary Elko, Bob Kubli, Jim West, and Rich Hochron for all their help in setting up the experimental apparatus and guidance in measurement techniques.

APPENDIX: IMPLEMENTATION OF THE FISCHELL–COKER ALGORITHM

This Appendix briefly describes how the Fischell–Coker talker direction finder algorithm[12] was implemented for obtaining the experimental results in this paper. The talker finder computes a cross correlation on specially processed versions of signals from two microphones. The processing (1) prewhitens so that correlation peaks will be sharp; (2) detects sudden increases in energy—a feature prominent in direct sound but not in reverberation; (3) emulates the precedence effect by excluding increases not preceded by 3 ms of weaker sound; (4) gain normalizes and quantizes to binary one or zero; and (5) computes the correlation as a simple interval histogram.

Inputs from each channel are first differenced, then sent to a Hilbert envelope detector. Fischell and Coker[12] refers to a bandpass filter for the analog front end. For our implementation, the bandpass filter is realized as the combination of the antialiasing sampling filter and a first difference operation (preemphasis). We used 16-kHz sampling, so the processing band was roughly 4–8 kHz for all of the experiments in this paper. The Hilbert envelope detector employs digital allpass filters with $90° \pm 0.2°$ relative phase from about 200 Hz to 7800 Hz. Following the Hilbert envelope operation is a stage equivalent to a diode-capacitor peak detector. A lag factor is chosen to produce a drop of 50%–80% in one pitch period, i.e., a number on the order of 0.97. Some additional noise/reverberation immunity is gained by repeating the process, using the derivative of the previous envelope as the input.

Event detection emulates the precedence effect by responding to increases in energy specifically preceded by a 3-ms interval with no increase. Processed signals are zero most of the time and occasionally of value one. Therefore, the cross correlation becomes identical to an interval histogram: If, for example, a left event follows n samples later than a right event, increment cell n of the histogram/cross correlation. We used four overlapping 400-ms windows and accepted a new target if its position accumulated a count greater than a user-defined threshold.

[1] H. F. Silverman and S. E. Kirtman, ''A two-stage algorithm for determining talker location from linear microphone array data,'' Comput. Speech Lang. **6**, 129–152 (1992).

[2] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, ''A closed-form location estimator for use with room environment microphone arrays,'' IEEE Trans. Speech Audio Process. **5**, 45–50 (1997).

[3] D. V. Rabinkin, R. J. Renomeron, A. Dahl, J. C. French, and J. L. Flanagan, ''A DSP implementation of source location using microphone arrays,'' Proc. SPIE **2846**, 88–99 (1996).

[4] B. Champagne, S. Bedard, and A. Stephenne, ''Performance of time-delay estimation in the presence of room reverberation,'' IEEE Trans. Speech Audio Process. **4**, 148–152 (1996).

[5] A. Stephenne and B. Champagne, ''Cepstral prefiltering for time delay estimation in reverberant environments,'' in *Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing* (IEEE, New York, 1995), pp. 3055–3058.

[6] M. Bodden, ''Modeling human sound-source localization and the cocktail-party-effect,'' Acta Acoustica **1**, 43–55 (1993).

[7] B. Kollmeier and R. Koch, ''Speech enhancement based on physiological and psychoacoustical models of modulation perception and binaural interaction,'' J. Acoust. Soc. Am. **95**, 1593–1602 (1994).

[8] W. Lindemann, ''Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals,'' J. Acoust. Soc. Am. **80**, 1608–1622 (1986).

[9] W. Gaik, ''Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling,'' J. Acoust. Soc. Am. **94**, 98–110 (1993).

[10] W. Lindemann, ''Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wave front,'' J. Acoust. Soc. Am. **80**, 1623–1630 (1986).

[11] A. G. Dabak and D. H. Johnson, ''Function-based modeling of binaural processing: Level and time cues,'' J. Acoust. Soc. Am. **94**, 2604–2616 (1993).

[12] D. R. Fischell and C. H. Coker, ''A speech direction finder,'' in *Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing* (IEEE, New York, 1984), pp. 19.8.1–19.8.4.

[13] W. C. Ward, G. W. Elko, R. A. Kubli, and W. C. McDougald, ''The new varechoic chamber at AT&T Bell Labs,'' in *Proceedings of the Wallance Clement Sabine Centennial Symposium* (Acoustical Society of America, Woodbury, NY, 1994), pp. 343–346.

Audible circuit noise in hearing aid amplifiers

Jeremy Agnew
Director of Product Development, Starkey Laboratories, Inc., Eden Prairie, Minnesota 55344

(Received 27 January 1997; accepted for publication 23 July 1997)

Audible electronic circuit noise generated within a hearing aid is distracting to a listener in quiet situations and, if the noise level is high enough, may cause listener irritation and rejection of the hearing aid. Thus for hearing aid specification and fitting purposes, it is useful to know the acoustic levels at which this internal noise may become audible and also at which it may become objectionable. For hearing aid amplifier circuit specification and design purposes, it is useful to know the same levels in electrical terms. This paper reports on a study that used an amplifier with no acoustic input and a hearing aid receiver output to simulate internally generated hearing aid circuit noise. Results are reported for testing eight subjects with high-frequency hearing loss for the perceived acoustic and electrical levels at which internal circuit noise became both audible and objectionable. © *1997 Acoustical Society of America.* [S0001-4966(97)01711-6]

PACS numbers: 43.58.Vb, 43.66.Ts [SLE]

INTRODUCTION

Internal amplifier noise generated within an audio system has always been looked upon as objectionable, because it adds undesired coloration to the reproduced sound. Similarly, a wearer of a hearing aid may object to audible circuit noise generated by the amplifier circuit. The listener's degree of awareness of this internal circuit noise depends on the listening situation. In many listening environments, existing ambient background noise is enough to mask out any perception of circuit noise. However, when listening in a quiet room or when listening to quiet passages of music, the relative perception for the listener can be that the internal level of the noise is quite loud. This study was prompted by comments from some wearers of hearing aids that, when listening under quiet conditions, they were able to hear the internal noise generated within their aids.

In recent years, the measurement and understanding of noise levels in hearing aids have increased in importance with advances in the technology used for hearing aid integrated circuit design and manufacture. Modern integrated circuits fabricated with complementary metal oxide semiconductor (CMOS) integrated circuit technology typically have three to ten times higher equivalent input noise than does the same circuit architecture implemented with low-noise bipolar devices (Gregorian and Temes, 1986). These higher internal noise levels have become more of an issue for both manufacturers and users, since internal circuit noise at these higher levels may be audible and become a source of irritation for the user.

If internal noise occurs at a high enough level, it may be disruptive to speech communication by direct acoustic masking of the desired signal (Jirsa and Norris, 1982; Moore, 1986; Teder, 1993). Another problem is the creation of undesired intermodulation products in the perceived sound, which can act as a source of masking (Agnew, 1988). Fielder (1985) has stated that noise signals with energy concentrated in the region from 1 to 3 kHz tend to produce the most effective masking.

Even if internal hearing aid noise does not occur at a level high enough to disrupt communication, many listeners object to its presence as a distracting artifact when in a quiet listening situation. Because it is not possible to realize a totally noiseless amplifier in practice, it is useful to know the level at which amplifier noise becomes audible, and the level at which it becomes objectionable or distracting to the listener. The purpose of the study reported here was to obtain measurements of the acoustic and electrical levels of internal noise in a hearing aid that were perceived by hearing-impaired listeners to be audible, and the levels that were perceived to be objectionable.

The problem of specifying levels of noise performance in a hearing aid becomes complex when considering that the end user has some degree of hearing impairment. Low levels of noise that are audible to a listener with normal hearing may be inaudible to a listener with a hearing impairment if the sound-pressure levels are below auditory threshold. However, as the amount of internal noise increases, the level eventually surpasses the listener's depressed hearing threshold and becomes audible. At some point it will become distracting and objectionable.

A simplified diagram of the pathway of hearing aid noise to the auditory system is shown in Fig. 1. The top row of icons, labeled ''signal path,'' shows the progression of circuit noise from its generation in the circuit, through the amplifier and receiver[1] to the ear, then its passage through the auditory system. In this simplified diagram, the noise is idealistically depicted as being created by a single noise generator and then amplified by a noiseless amplifier. In practice, this is a fairly accurate representation because the dominant noise source in a hearing aid is usually the microphone.

The primary sources of noise in the amplifier are the active devices, such as transistors and integrated circuits; however, even resistors can contribute significantly to the total electronic noise. There are three common types of noise generated within integrated amplifiers. These are white noise produced by thermal and shot mechanisms, burst noise, and

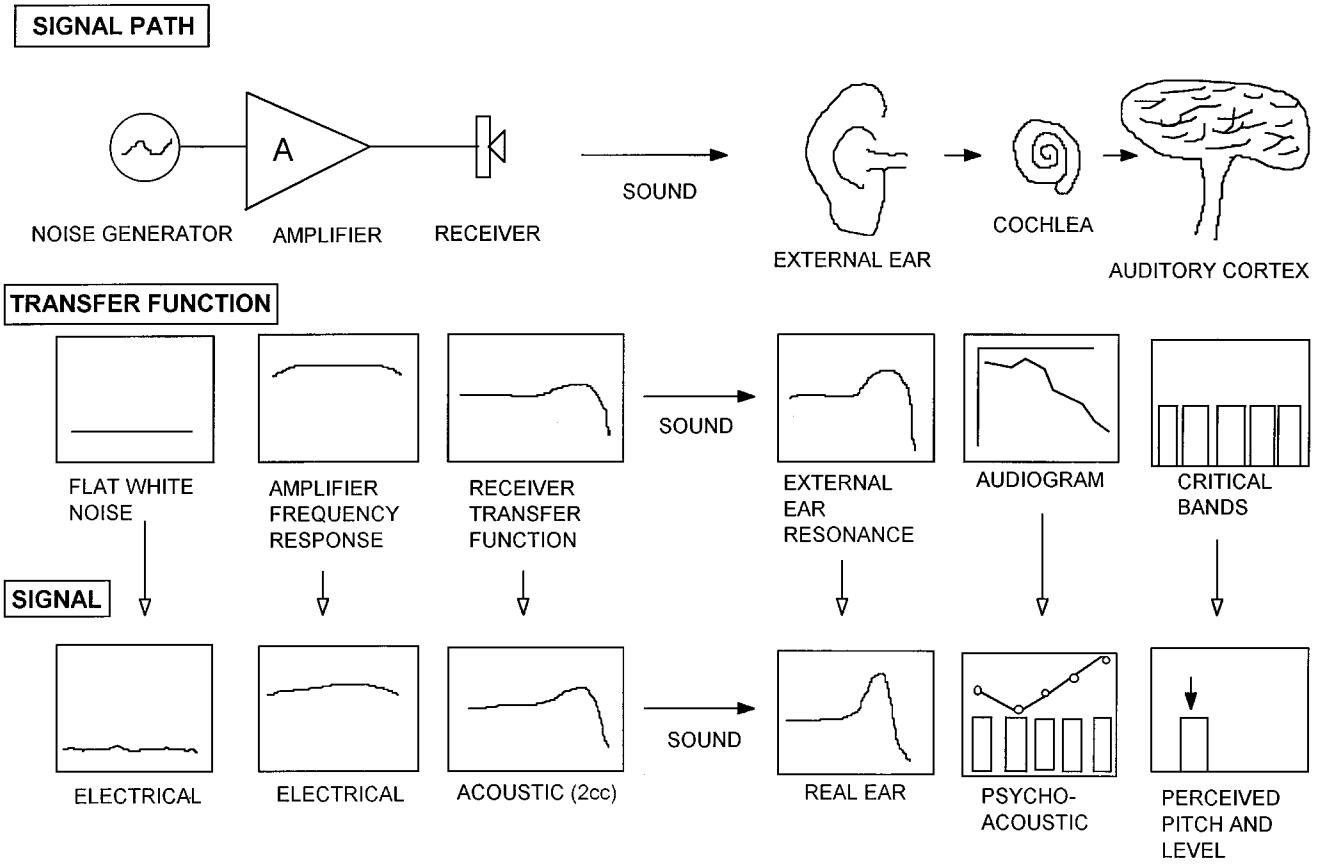


FIG. 1. Simplified schematic representation of the perception of internally generated hearing aid noise traveling from the amplifier noise source to the auditory cortex. The first row of icons represents the signal path; the second row of rectangles represents the transfer function of the block above it in the signal path; and the third row of rectangles represents the spectral shape of the noise, as acted upon by the transfer function of the block.

flicker noise (Fish, 1994). Thermal noise and shot noise have different noise-producing mechanisms, but both result in white noise, which sounds like a constant low-level hiss in the output of the hearing aid. Burst noise (also known as popcorn noise) adds intermittent popping or crackling sounds. This is a particularly objectionable type of noise, because of its intermittent nature and abrupt onset. Flicker noise is a random noise that is inversely related to frequency and, because of this, is also commonly known as $1/f$ noise (''one-over-eff'' noise). Flicker noise produces its highest output level at low frequencies, often within the lower end of the audio frequency amplifying range. Since most sensorineural hearing losses result in hearing at low frequencies within limits considered to be normal, the low-frequency content of $1/f$ noise may be high enough to become audible to the wearer of a hearing aid.

As seen in the second and third circuit blocks in the top row of icons in Fig. 1, the noise is amplified, then converted by the receiver to a sound in the ear canal. The sound passes from the external ear, through the middle ear, to the cochlea, and then to the auditory cortex. As the noise passes through this electronic, acoustic, and biological signal pathway, it is modified by various transfer functions. The amplifier frequency response is shown in this diagram as flat, except in the extreme low- and high-frequency regions. In practice, there would be frequency response shaping in order to compensate for the specific hearing loss of the listener. However, even if $1/f$ noise is the dominant source, intentional low-frequency attenuation by the circuit designer may eliminate any low-frequency emphasis from the overall output.

Spectral shaping of the noise is strongly affected by the transfer function of the receiver, which typically acts as a low-pass filter with a peak around the roll-off frequency. The amplified noise is spectrally modified by the internal frequency response characteristics of the hearing aid, and dominated by the acoustic frequency response and passband of the receiver. Thus the long-term spectrum of the internal noise of a hearing aid has essentially the same shape as the acoustic frequency response. An example of a typical long-term spectrum of internal hearing aid noise is shown in Fig. 2.

As the sound passes into the cochlea, perception is modified by the hearing loss of the listener, exemplified in Fig. 1 as an audiogram with a high-frequency sensorineural hearing loss, which results in a lowered sensitivity to high-frequency sounds. Thus the low-frequency components of the noise may be audible, whereas the high-frequency components fall below the listener's threshold of hearing. Finally, the sound is typically perceived by the listener as though presented through a parallel bank of overlapping level-dependent filters with a bandwidth equivalent to critical bands (Fletcher, 1940). Critical bands are roughly one-third-octave in width over a fairly wide frequency range (Moore, 1982), though this approximation of critical bands by one-

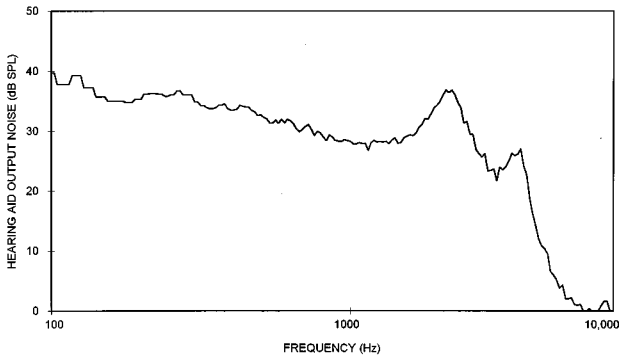


FIG. 2. Typical long-term spectrum of the internal noise generated by a hearing aid.

third-octave filters is only acceptable above about 300 Hz (Zwicker and Fastl, 1990).

Very little has appeared in the literature concerning the audibility of internal hearing aid circuit noise. The ANSI hearing aid measurement standard describes a standardized method for calculating a single-figure equivalent input noise level for a hearing aid and in Annex C, shows an optional method of displaying hearing aid output noise in one-third-octave bands (ANSI, 1996). However, since this is a quality control standard, no attempt is made to relate this calculated noise figure to perception.

The sound pressure level of hearing aid output noise should be distinguished as having a different meaning for perception than the equivalent input noise (EIN) figure. The EIN figure is calculated by subtracting the high-frequency average gain of the hearing aid from the output noise level (ANSI, 1996). Thus, though a high gain hearing aid will have a higher level of output noise than a low gain aid, this calculation removes the influence of the gain to allow a comparison of hearing aids. Thus it is possible to have a good EIN figure, but to still have a high absolute level of output noise. The hearing aid wearer hears the absolute level of the output noise, not the EIN.

Stuart (1994) has presented a review of the underlying psychoacoustic mechanisms of noise detection and threshold phenomena, and contrasted various measures of human auditory frequency selectivity. The discussion related to normal hearing and a method of modeling the human auditory process, and was not extended to listeners with hearing impairment. Killion (1976) has compared the noise level of a subminiature research hearing aid microphone to the internal noise of the human ear; however, no attempt was made to study the audibility of microphone noise.

Dillon and MacRae (1984), and MacRae and Dillon (1986, 1996) have described criteria for maximum equivalent input noise considered acceptable when listening to a speech signal with a long-term rms level of 65 dB SPL, based on an acceptable signal-to-noise ratio (SNR) in one-third-octave bands. However, these criteria relate to signal-to-noise levels, not to noise audibility in the situation where internal circuit noise from the hearing aid alone is the stimulus.

A study of the audibility of internal hearing aid circuit noise by Agnew (1996) related internal hearing aid circuit noise to the hearing thresholds of eight listeners with moderate hearing losses. While listening to a hearing aid that produced a variable level of internal circuit noise, each participant in the study was asked to indicate the level at which internal circuit noise just became audible. Following this, the spectrum of the noise was measured in one-third-octave bands at the chosen setting, and the audiometric thresholds of the listener were converted to the corresponding SPL values. Superimposing the two sets of data on the same graph showed that the internal hearing aid noise first became audible when the level in one of the one-third-octave bands coincided with the audiometric threshold of the listener at some frequency, typically at the frequency of greatest sensitivity of the residual hearing.

Egan and Meyer (1950), and de Boer (1962) speculated that the area on the basilar membrane that has the highest SNR determines the perception of the pitch of a sound. Since the highest SNR should occur at the frequency of the most sensitive hearing, it would be expected that this would result in a pitch-match at this frequency. This concept is also reasonable in light of recent loudness models based on excitation patterns (Zwicker and Fastl, 1990; Humes, 1994; Moore, 1995).

I. METHOD

A. Subject selection

Eight hearing-impaired individuals were recruited as subjects. All were experienced hearing aid users. General selection criteria included a normal tympanogram, normal otoscopic examination, and a moderate hearing loss, particularly in the high frequencies. Individual audiograms for the eight subjects are shown in Fig. 3.

Subjects with moderate hearing losses were selected for this study as being representative of a group of individuals with common hearing losses who might object to high levels of internal circuit noise in hearing aids, and who might benefit from a reduction of internal hearing aid noise. Mild hearing losses were excluded from this study because these individuals were assumed to be relatively close to normal-hearing individuals in their perception of audible noise levels. Subjects with severe hearing losses were also excluded from this study, since their losses are usually such that low levels of internal noise in hearing aids are not audible to them. Experienced hearing aid wearers were chosen because they were already aware of circuit noise and had experience with the type of sound that they would be hearing.

B. Test hearing aid

A hearing aid amplifier circuit was constructed in a test box with a connector at the output to allow the attachment of different hearing aid receiver modules. The circuit architecture was that of the conceptual hearing aid shown in Fig. 1, and was composed of a noise source having variable amplitude connected to the input of a low-noise amplifier driving a hearing aid receiver.

The test amplifier was constructed with Analog Devices OP-270 low-noise precision audio operational amplifiers, in

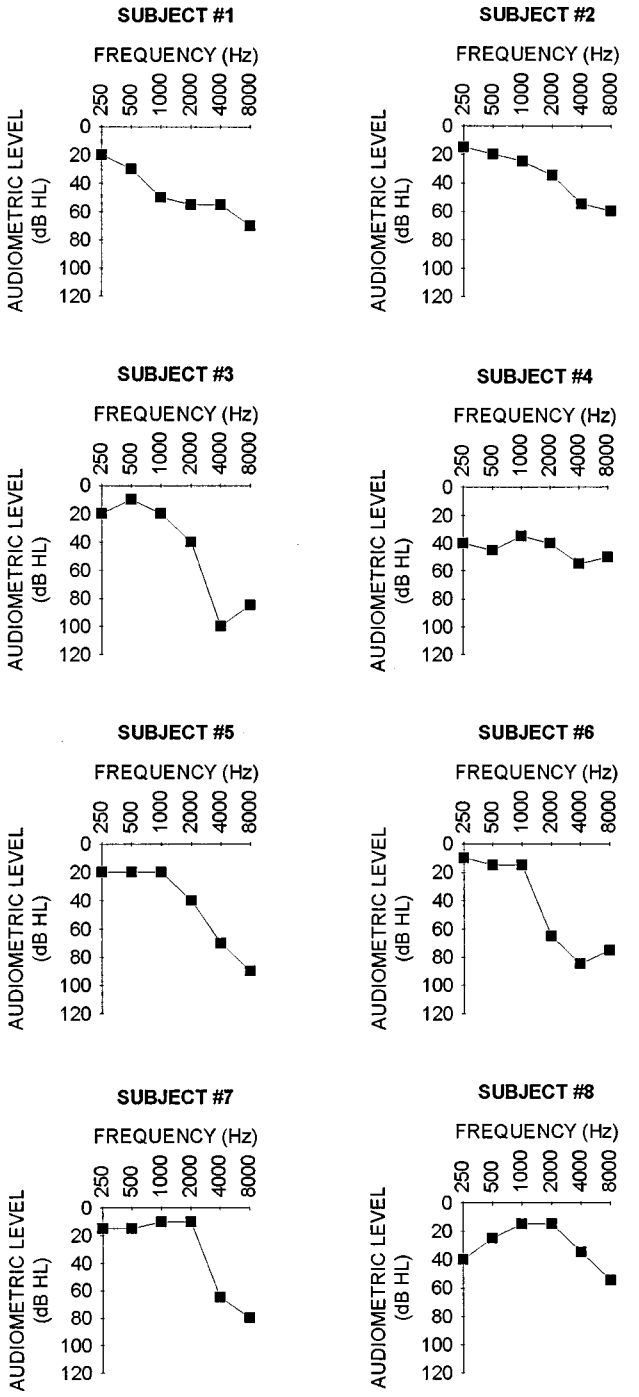


FIG. 3. Right ear audiograms for the eight subjects expressed in decibels HL.

order to produce inherent system noise levels below those currently achievable with off-the-shelf hearing aid amplifiers. To control the level of the noise accurately, and to include the ability to test with noise levels higher and lower than those typically found in hearing aid amplifiers, the test circuit was built with a generator that could be used to produce a variable level of electronic noise. The output of the generator was connected to a switched attenuator that varied between minimum and maximum output with a step size of 3 dB. An attenuator with a fixed step size was chosen, rather than a continuously variable control, in order to reproduce exactly each subject's setting for acoustic measurement following the psychoacoustic test session. The output transducer for each subject was a Knowles model EP4107 hearing aid receiver, built into a custom in-the-ear (ITE) hearing aid module made for the right ear of each subject.

C. Test and presentation method

The test session was performed with each subject seated in a double-walled audiometric test booth. Testing of noise levels was initiated with the noise generator control set to the maximum attenuation position.

Two test conditions were explored:

(1) The subject was instructed to advance the output noise control to the position where the level of the noise was considered, in their perception, to be just audible. The subject's response was recorded as the setting that was consistent for 3 out of 5 trials.
(2) The subject was then instructed to advance the attenuator to the position where the subject felt that the noise was objectionable. ''Objectionable'' was defined for the purposes of this study as the level of noise that the subject felt would be unpleasant to listen to for an extended period of time. As before, the subject's response was recorded as the setting that was consistent for 3 out of 5 trials.

D. Electroacoustic measurements

Following the perceptual tests, the spectra and levels of the noise were measured electrically and acoustically at the setting corresponding to each subject's perceived audible and objectionable levels.

The electrical levels were recorded as the rms voltage across the output transducer. This measurement was performed using a Bruel & Kjaer 2032 dual-channel signal analyzer. One hundred samples were exponentially averaged to derive the final recorded value.

The acoustic measurements were made using the same dual-channel signal analyzer connected to a Bruel & Kjaer type 2619 preamplifier and 4134 $\frac{1}{2}$-in. microphone with a Knowles Electronics' Zwislocki coupler. The receiver modules were sealed into the coupler at a uniform depth of 1 cm, in order to approximate a consistent insertion depth into the subjects' ears. While it was realized that this was not a direct measure of the sound-pressure level generated in each individual subject's ear canal, this method was chosen as a standard and repeatable coupling method, under the assumption that resulting data would relate to the tests normally used in the design and quality control of hearing aids. Although a 2cc-coupler is used for hearing aid quality control measurements (ANSI, 1996), the conversion back and forth to a Zwislocki coupler is a simple mathematical conversion (Killion and Revit, 1993).

II. RESULTS AND DISCUSSION

Figure 4 superimposes the individual subject audiograms, converted to decibels SPL using conventional audio

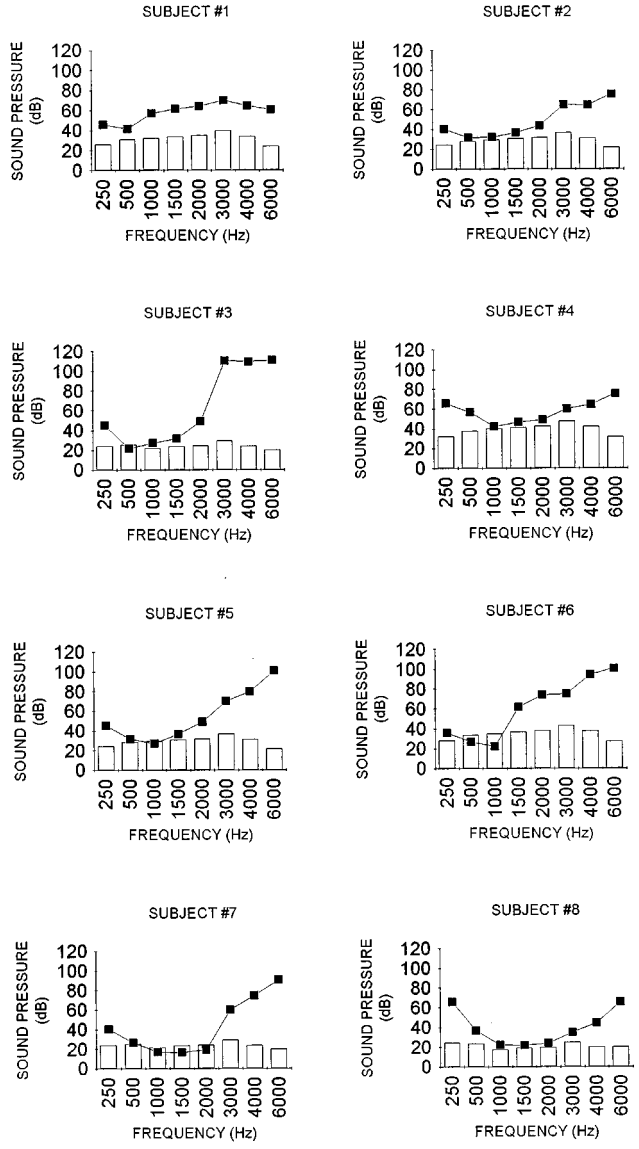


FIG. 4. Graphs by subject of the one-third octave noise figures for the hearing aid circuit adjusted to the subject's perception of just-audible output noise, superimposed on the subject's audiometric loss converted to decibels SPL. Circuit noise is plotted as the vertical bars; audiometric loss is plotted as the squares with connecting lines.

metric correction figures from ANSI (1989), on the one-third-octave values for the just-audible noise at each audiometric frequency. It can be seen that the noise typically became audible when a one-third-octave value from the internal noise measurement of the hearing aid extended up to the lowest value of the subject's SPL audiogram.

For the graphs of subjects #6 and #7 in Fig. 4, there is a slight overlap of audiogram and one-third-octave noise graphs, instead of an intersection of the two plots. Several possible sources of error could have contributed to this result. First, it is possible to speculate that these two subjects interpreted the instruction ''just-audible'' to be a slightly higher sound level than did the other subjects. A second possibility is that clinical audiograms are typically considered accurate to ± 5 dB, and this accuracy may have affected the results. A third possible source of error is that the noise source could only be varied in steps of 3 dB. This, combined with the tolerance of the threshold measurements, may not have resulted in a fine enough resolution to pinpoint the exact threshold of audibility of the noise. A fourth possibility is that using a standard insertion depth for the electroacoustic coupler measurements may have caused the coupler measurement to be different than the actual sound pressure in the subject's ear canal.

One or all of these possibilities may have also affected the accuracy of the graphs for subject #1, for whom the graphs do not quite intersect. Another possibility for this subject is that the frequency of the most sensitive hearing could have occurred at a frequency in between the measured standard audiometric frequencies.

Table I shows data for each subject. The column labeled *audible* contains the one-third-octave band sound-pressure levels for the noise considered by each individual to be just audible at each audiometric frequency. The column labeled *objectionable* contains the values for the acoustic noise levels considered to be objectionable by each individual. The last column, labeled *audiogram*, is the audiometric data converted from HL to SPL values. Bold italic type identifies the audiometric frequency of each subject's most sensitive hearing, and corresponds to the frequency at which the audiogram intersected the one-third-octave noise values, as shown graphically in Fig. 4.

For each subject, Table I provides the one-third-octave sound pressure level at which the noise was identified to be audible. Thus, for subject #8, if the one-third-octave value of the noise for any hearing aid would exceed 19.1 dB SPL at 1500 Hz, the subject would probably identify the noise as being audible. The listing of objectionable one-third-octave noise values can be used similarly. For example, if the one-third-octave value of the noise of a hearing aid were to exceed 33.7 dB at 1500 Hz for subject #8, this subject would probably identify the noise as being objectionable.

In Table II, the column labeled *frequency* contains the frequency of most sensitive hearing, by subject, along with the corresponding one-third-octave sound-pressure level for the audible and objectionable perceptions. The next column is the difference (rounded off) between the audible and objectionable values. This difference ranged between 4 dB greater than the audible value for subject #7, to 15 dB greater than the audible value for subject #8. The reason for this wide range of differences is not apparent and no correlation was seen between these differences and the other data. It is possible to speculate that this could be due to a wide variety of causes, from individual differences in pathology of the hearing loss, to individual perceptions of how loud the sound had to be to be considered objectionable.

Table II also shows the electrical value of the signal across the receiver corresponding to the audible acoustic value, and the electrical value corresponding to the objectionable acoustic value. Both of these were obtained by measuring the one-third-octave value of the electrical signal across the receiver at the corresponding acoustic level. These values would be useful to a circuit designer who is performing circuit simulations and who needs to know the values of electrical output from an amplifier circuit that would be con-

TABLE I. One-third octave data by subject for audiometric frequencies.

Subject	Freq. (Hz)	Audible dB SPL[b]	Object-ionable dB SPL	Audiogram dB SPL	Subject	Freq. (Hz)	Audible dB SPL	Object-ionable dB SPL	Audiogram dB SPL
#1	250	26.1	34.3	45.5	#5	250	24.5	28.1	45.5
	500[a]	***30.7***	***39.0***	***41.5***		500	28.2	33.2	31.5
	1000	32.3	41.6	57.0		***1000***	***29.2***	***35.0***	***27.0***
	1500	33.7	43.5	61.5		1500	30.5	36.8	36.5
	2000	34.9	44.2	64.0		2000	31.8	38.1	49.0
	3000	39.9	49.2	70.0		3000	36.6	43.0	70.0
	4000	34.2	44.1	64.5		4000	31.1	37.4	79.5
	6000	24.2	33.6	60.5		6000	21.4	27.2	100.5
#2	250	24.4	31.9	40.5	#6	250	28.1	36.4	35.5
	500	***28.2***	***37.6***	***31.5***		500	33.3	42.3	26.5
	1000	29.2	39.8	32.0		***1000***	***34.9***	***44.8***	***22.0***
	1500	30.5	41.3	36.5		1500	36.8	46.4	61.5
	2000	31.8	42.7	44.0		2000	38.1	47.7	74.0
	3000	36.6	47.6	65.0		3000	43.0	52.5	75.0
	4000	31.1	42.1	64.5		4000	37.4	46.8	94.5
	6000	21.4	31.7	75.5		6000	27.2	36.6	100.5
#3	250	23.9	34.3	45.5	#7	250	23.9	23.7	40.5
	500	***25.3***	***39.0***	***21.5***		500	25.3	25.9	26.5
	1000	21.5	41.6	27.0		1000	21.5	25.6	17.0
	1500	23.3	43.5	31.5		***1500***	***23.3***	***27.3***	***16.5***
	2000	24.5	44.3	49.0		2000	24.5	28.7	19.0
	3000	29.2	49.2	110.0		3000	29.1	33.6	60.0
	4000	23.9	44.1	109.5		4000	23.9	28.1	74.5
	6000	19.6	33.6	110.5		6000	19.6	19.8	90.5
#4	250	31.9	36.4	65.5	#8	250	24.3	26.1	65.5
	500	37.6	42.3	56.5		500	23.6	30.7	36.5
	1000	***39.8***	***44.7***	***42.0***		1000	17.1	32.3	22.0
	1500	41.3	46.4	46.5		***1500***	***19.1***	***33.7***	***21.5***
	2000	42.7	47.7	49.0		2000	19.5	34.9	24.0
	3000	47.7	52.5	60.0		3000	24.8	39.9	35.0
	4000	42.1	46.8	64.5		4000	19.8	34.2	44.5
	6000	31.7	36.6	75.5		6000	19.6	24.2	65.5

[a]The frequency in ***bold italic type*** for each subject indicates the frequency of the most sensitive hearing for that subject, corresponding to the audiometric curves in Fig. 4.
[b]SPL data were measured in a Zwislocki coupler with the same signal voltage impressed across the receiver.

sidered to be audible and/or objectionable. The designer might then modify the electrical circuit design to vary these values in order to modify the particular listener's perception of internal noise. It is important to note, however, that these electrical values are only valid for the model of receiver used for this testing, which was Knowles Electronics model EP4107. To apply these data to other receivers, an appropriate transfer function from electrical to acoustic values for

TABLE II. Summary of acoustic and electrical audible and objectionable noise levels, measured in a Zwislocki coupler to produce approximate eardrum SPL.

Subject	Frequency (Hz)	Audible acoustic level (dB SPL)	Objectionable acoustic level (dB SPL)	Difference between audible and objectionable (dB values rounded)[a]	Audible electrical level (with Knowles EP4107 receiver) (microvolts rms)[b]	Objectionable electrical level (with Knowles EP4107 receiver) (microvolts rms)[b]
#1	500	30.7	39.0	8	117	352
#2	500	28.2	37.6	9	80	285
#3	500	25.3	39.0	14	35	352
#4	1000	39.8	44.7	5	411	721
#5	1000	29.2	35.0	6	117	234
#6	1000	34.9	44.8	10	234	721
#7	1500	23.3	27.3	4	63	104
#8	1500	19.1	33.7	15	38	220

[a]Mean difference=8.8 dB.
[b]Valid for EP4107 receiver only.

that particular model of receiver would have to be measured or calculated.

III. CONCLUSIONS

This study showed that, for a group of eight subjects with moderate sensorineural hearing loss, the threshold of audibility of internal hearing aid circuit noise corresponded individually to the noise level where the one-third-octave value of the noise at the frequency of the most sensitive hearing intersected the value of the threshold of hearing in SPL. The acoustic values corresponding to each subject's perception of audible noise levels were presented. The noise level was advanced until each subject felt that the noise level was objectionable, according to their perception and interpretation of the descriptor *objectionable*.

ACKNOWLEDGMENTS

The author would like to thank Jerry Wahl of the research department of Starkey Laboratories for the design and construction of the test amplifier, and Lisa Potts, research audiologist for the Division of Adult Audiology of Washington University, for recruiting and testing the subjects.

[1]In hearing aid terminology the word *receiver* is used to denote the electrical to acoustic output transducer. In other areas of acoustics this is usually called the *earphone*.

Agnew, J. (**1988**). ''Hearing instrument distortion: What does it mean for the listener?,'' Hear. Instr. **39**, 10, 12, 14, 16, 20, 61.

Agnew, J. (**1996**). ''Perception of internally-generated noise in hearing amplification,'' J. Am. Acad. Audiol. **7**, 296–303.

American National Standards Institute (ANSI) (**1989**). ANSI S3.6-1989, ''American National Standard Specification for Audiometers'' (Acoustical Society of America, New York).

American National Standards Institute (ANSI) (**1996**). ANSI S3.22-1996, ''American National Standard Specification of Hearing Aid Characteristics'' (Acoustical Society of America, New York).

de Boer, E. (**1962**). ''Note on the critical bandwidth,'' J. Acoust. Soc. Am. **34**, 985–986.

Dillon, H., and Macrae, J. (**1984**). ''Derivation of design specifications for hearing aids,'' National Acoustics Laboratories Report Number 102 (Australian Government Publishing Service, Canberra, Australia).

Egan, J. P., and Meyer, D. R. (**1950**). ''Changes in pitch of tones of low frequency as a function of the pattern of excitation produced by a band of noise,'' J. Acoust. Soc. Am. **22**, 827–833.

Fielder, I. D. (**1985**). ''The audibility of modulation noise in floating-point conversion systems,'' J. Audio Eng. Soc. **33**, 770–781.

Fish, P. J. (**1994**). *Electronic Noise and Low Noise Design* (McGraw-Hill, New York).

Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65.

Gregorian, R., and Temes, G. C. (**1986**). *Analog MOS Integrated Circuits for Signal Processing* (Wiley, New York).

Humes, L. E. (**1994**). ''Psychoacoustic considerations in clinical audiology,'' in *Handbook of Clinical Audiology*, edited by J. Katz (Williams & Wilkins, Baltimore), 4th ed.

Jirsa, R. E., and Norris, T. W. (**1982**). ''Effects of intermodulation distortion on speech intelligibility,'' Ear Hear. **3**, 251–256.

Killion, M. C. (**1976**). ''Noise of ears and microphones,'' J. Acoust. Soc. Am. **59**, 424–433.

Killion, M. C., and Revit, L. J. (**1993**). ''CORFIG and GIFROC: Real ear to coupler and back,'' in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston), 2nd ed.

Macrae, J., and Dillon, H. (**1986**). ''Updated performance requirements for hearing aids,'' National Acoustics Laboratories Report No. 109 (Australian Government Publishing Service, Canberra, Australia).

Macrae, J., and Dillon, H. (**1996**). ''An equivalent input noise level criterion for hearing aids,'' J. Rehabil. Res. Dev. **33**, 355–362.

Moore, B. C. J. (**1982**). *An Introduction to the Psychology of Hearing* (Academic, London).

Moore, B. C. J. (**1986**). *Frequency Selectivity in Hearing* (Academic, London).

Moore, B. C. J. (**1995**). *Perceptual Consequences of Cochlear Damage* (Oxford U.P., Oxford).

Stuart, J. R. (**1994**). ''Noise: Methods for estimating detectability and threshold,'' J. Audio Eng. Soc. **42**, 124–139.

Teder, H. (**1993**). ''Compression in the time domain,'' Am. J. Audiol. **2**, 41–46.

Zwicker, E., and Fastl, H. (**1990**). *Psychoacoustics, Facts and Models* (Springer-Verlag, Berlin).

Free-mass reciprocity calibration

Thomas B. Gabrielson[a)]
NAWC Aircraft Division, Code 45544, MS 07, Warminster, Pennsylvania 18974-0591

(Received 21 January 1997; accepted for publication 31 July 1997)

A simple method for reciprocity calibration of accelerometers and velocity sensors is presented. The useful frequency range is from about 1 Hz to a kilohertz or more and the method is capable of subpercent accuracy. Calibrations with excitation levels from 1 to 300 mm/s^2 (0.1 mg to 0.03 g) were produced in prototype fixtures. The device to be calibrated and two geophones are used in the calibration fixture. Since the geophones are both reciprocal, a direct test of reciprocity can be done prior to the calibration to evaluate the apparatus. In contrast to standard techniques, the method requires only one relatively inexpensive mechanical setup. [S0001-4966(97)05111-4]

PACS numbers: 43.58.Vb, 43.40.Yq, 43.38.Ar [SLE]

INTRODUCTION

Reciprocity calibration is often assumed to be a method restricted to occasional calibration of primary standards. There are, however, instances when reciprocity provides a means for convenient calibration of sensors on a routine basis.[1,2] Common methods for reciprocity calibration of vibration sensors involve three experimental setups and an auxiliary exciter and are capable of good precision well into the kilohertz region.[3–9] An alternate technique, presented here, uses only a single mechanical setup with inexpensive geophones as the excitation and reciprocal transducers. While the technique is more difficult to extend above 1 kHz than the usual technique, at lower frequencies it is considerably easier, requires no expensive hardware, and is still capable of subpercent accuracy. In addition to being useful for routine calibration, the simplicity of the setup makes this technique ideal for demonstrating reciprocity calibration in a teaching laboratory.

A reciprocal transducer (or, more generally, a reciprocal device or network) is a two-port device for which the ratio of potential produced at one set of terminals to flow into the other set is identical when driven at either terminal. For example, for a network of linear resistors, capacitors, and inductors, the ratio of voltage produced at terminal 2 to drive current at terminal 1 is identical to the ratio of voltage produced at terminal 1 to drive current at terminal 2. Reciprocity does not depend on a device being lossless; purely resistive linear networks are reciprocal. Typically, calibration requires three measurements with three transducers: a source, S, a receiver, R, and a reciprocal transducer, T. S is driven and the output of R measured; S is driven and the output of T is measured; and T is driven and the output of R is measured. There are four unknowns: the transmitting responses of S and T and the receiving responses of R and T. With the three measurements and the one relationship given by reciprocity (for the reciprocal device, T), the four responses are determined. The key to developing useful reciprocity calibration techniques is to arrange the measurement so that only quantities that are easy to measure accurately (masses, voltage ratios, resistance, distances, for example) are required. In the procedure discussed below, the transducer responses (voltage out for velocity in, and force out for current in) are determined using absolute measurements of only resistance and mass. (Reciprocity does not obviate the need for absolute measurements!) A freely suspended mass is used to transform forces to velocities and the reciprocity relationship is used to remove the forces and velocities from the equation for the transducer response.

I. BASIC PROCEDURE

The technique involves clamping two geophones and the sensor to be calibrated between two metal bars or channels and suspending the assembly as a pendulum (see Fig. 1). Alternately, the assembly can be oriented vertically and suspended by surgical tubing. Above the pendulum resonance, the pendulum acts as a nearly pure mass in converting an applied force to a velocity. One of the geophones is used as a source (S), another geophone is used as both a source and receiver (T), and the transducer to be calibrated is used as a receiver (R). Only unit T needs to be reciprocal. The technique does not depend on S and T being geophones but geophones are convenient because they are reciprocal and are also capable of sufficient inertial reaction to drive the pendulum assembly when clamped to that assembly.

The receiving response α will be taken as the open-circuit voltage e per case velocity v (the normal definition for a geophone):

$$\alpha = \left(\frac{e}{v}\right)_{i=0}. \tag{1}$$

This response definition will also be used for the device to be calibrated even if it is not a geophone; once the response in terms of velocity has been found, it can easily be converted in terms of acceleration. The transmitting response β of a geophone is the ratio of force produced for a given input current for zero case velocity:

$$\beta = \left(\frac{F}{i}\right)_{v=0}. \tag{2}$$

[a)]Presently at the Applied Research Laboratory, The Pennsylvania State University, P.O. Box 30, State College, PA 16804.

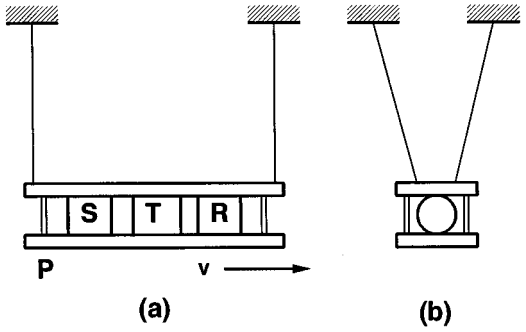


FIG. 1. Pendulum suspension for reciprocity calibration of geophones. (a) Side view and (b) end view of four-point suspension.

A third quantity, γ, is useful to incorporate the effects of the pendulum. This quantity is the ratio of pendulum velocity produced for a current into the geophone:

$$\gamma=\left(\frac{\nu}{i}\right)_P=\beta\,\frac{1}{Z_P}. \tag{3}$$

The subscript P denotes that this value refers to the geophone mounted on the pendulum fixture. Unlike α and β, the value of γ is specific to the experimental setup through the mechanical impedance, $Z_P(=F/\nu)$, of the apparatus.

The relationship between α_T and β_T (the subscripts indicate the relevant transducer) is obtained from reciprocity. If the pendulum can be treated as a lumped mass (a convenient but unnecessary simplification), then the connection of the geophone T to the pendulum P can be diagrammed as in Fig. 2. The mobility analogy is selected for the mechanical side: drawn in this manner, both T and P are reciprocal;[10] consequently, the ratio of the potential quantity (e or ν) at the receive terminal to the flow quantity (F or i) at the drive terminal is the same regardless of which terminals are selected for receive and drive:

$$\left(\frac{\nu}{i}\right)_{F=0}=\left(\frac{e}{F}\right)_{i=0}. \tag{4}$$

This conclusion can also be reached without invoking reciprocity by analysis of the full electromechanical equivalent circuit shown in Fig. 3. Define Z_P as the transfer impedance, F/ν', across the pendulum with T attached. Then, for zero applied force, the pendulum velocity, ν or ν', for a given current is

$$\nu=i\,\gamma_T \quad \text{or} \quad \left(\frac{\nu}{i}\right)_{F=0}=\gamma_T \tag{5}$$

and the open-circuit output of T for an applied force is

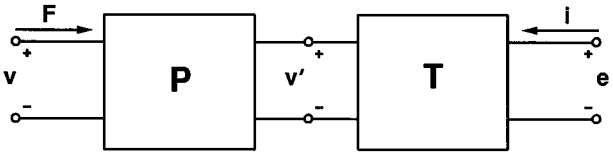


FIG. 2. Schematic representation of coupling between geophone (T) and pendulum (P) as linear, two-port networks.

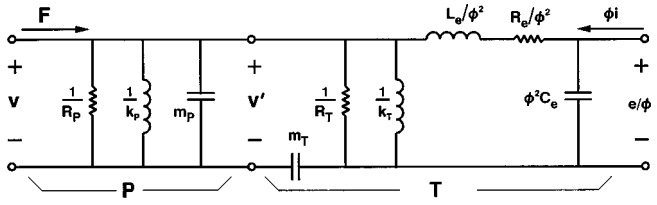


FIG. 3. Equivalent circuit for pendulum (P) with attached geophone (T).

$$e=\frac{F}{Z_P}\alpha_T \quad \text{or} \quad \left(\frac{e}{F}\right)_{i=0}=\frac{\alpha_T}{Z_P}. \tag{6}$$

From Eqs. (3), (4), and (6)

$$\alpha_T=\beta_T. \tag{7}$$

For the first measurement (measurement ST), the source, S, is driven by some current, i_{SA}, and the open-circuit output voltage, e_{TA}, of T is measured. Normally, the input current would be measured by the voltage drop, e_{SA}, across a reference resistor, R_0, in series with S (see Fig. 4). This measurement can be written as

$$e_{TA}=\frac{e_{SA}}{R_0}\gamma_{S,ST}\alpha_T=\frac{e_{SA}}{R_0 Z_{ST}}\;\beta_S\alpha_T. \tag{8}$$

The mechanical impedance, Z, is assumed to be different in each measurement, hence the subscripts. In order to make absolute measurement of voltages unnecessary, the measurement can also be expressed as the ratio of two voltages as follows:

$$[ST]\equiv\frac{e_{TA}}{e_{SA}}=\frac{\beta_S\alpha_T}{R_0 Z_{ST}}. \tag{9}$$

The ratio, denoted by the symbol $[ST]$, would also be the transfer function as measured by the two-channel spectrum analyzer in Fig. 4. If the analyzer has a swept-sine mode or random-noise excitation, the measurements can be made rapidly over wide bands.

The second measurement (SR) requires measurement of the drive current of S and the open-circuit output voltage of R:

$$[SR]\equiv\frac{e_{RB}}{e_{SB}}=\frac{\beta_S\alpha_R}{R_0 Z_{SR}} \tag{10}$$

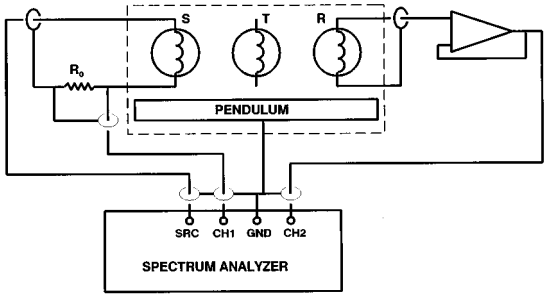


FIG. 4. Schematic of electrical connections for reciprocity calibration. The transducers, S, T, and R, are rigidly attached to the pendulum. The spectrum analyzer can be replaced by a lock-in amplifier or by a signal generator and an ac voltmeter. Not shown is an oscilloscope for monitoring the channel 1 and 2 signals.

while the third measurement (TR) requires driving T and measuring the output of R:

$$[TR] \equiv \frac{e_{RC}}{e_{TC}} = \frac{\beta_T \alpha_R}{R_0 Z_{TR}}. \tag{11}$$

Combining the results from the three measurements and Eq. (7) produces the receiving (or transmitting) response of T:

$$\alpha_T^2 = R_0 \frac{[ST] Z_{ST} [TR] Z_{TR}}{[SR] Z_{SR}}. \tag{12}$$

Notice that each of the bracketed factors in Eq. (12) is proportional to the product of the indicated responses [see Eqs. (9)–(11)]. For example, $[ST]$ is proportional to the product of β_S and α_T [see Eq. (4)]. This means that the correct form for Eq. (12) [and Eqs. (13) and (14) below] can be remembered by the mnemonic of ''cancelling'' letters between the bracketed factors.

In the same manner, the receiving response of R and the transmitting response of S can be determined:

$$\alpha_R^2 = R_0 \frac{[SR] Z_{SR} [TR] Z_{TR}}{[ST] Z_{ST}}, \tag{13}$$

$$\beta_S^2 = R_0 \frac{[ST] Z_{ST} [SR] Z_{SR}}{[TR] Z_{TR}}. \tag{14}$$

To make the calibration process practical, the apparatus must be designed so that an accurate determination of the mechanical impedances is possible. The freely suspended mass technique leads to a mechanical impedance that is dominated by the ordinary pendulum mass. The mass term is much larger than any of the other terms in the mechanical dynamics over several orders of magnitude in frequency. For many applications, the pendulum can be treated as a simple mass and the transfer impedance is

$$Z_P \approx j \omega m_P. \tag{15}$$

The pendulum mass m_P includes the masses of the three transducers on the pendulum. This approximation breaks down near the pendulum resonance frequency (easily contrived to be below 1 Hz) on the low end and at sufficiently high frequencies that the pendulum bar can no longer be considered a lumped mass. Also, above the resonance frequency of the geophone, the proof mass should not be included in m_P, just the case mass. This is a small and often negligible correction if accuracy of a few percent is adequate. Other corrections are considered in Sec. IV.

II. PRACTICAL IMPLEMENTATION

A very simple implementation adequate for calibration within a few percent to a few hundred hertz involves clamping the three transducers between two bars or channels as shown schematically in Fig. 1. For best results, the fixture should be constructed so that the axes of all of the transducers are aligned and the fixture is symmetric about that axis.

As a practical matter, a small amount of care in setting up the pendulum will make the measurements easier and more accurate. As shown in Fig. 1, a four-point suspension should be used. The parallelogram arrangement (side view) prevents pitchlike rotation of the pendulum bar, which would introduce a moment-of-inertia term into the transfer impedance. The triangular arrangement (end view) suppresses lateral oscillation. While there are in-band resonances in the suspension lines, they are not well coupled to the pendulum mass and do not appear to affect the pendulum transfer impedance. The leads from the geophones should be thin, stranded wire, loosely coiled and suspended near the pendulum to minimize the effect on the pendulum dynamics. If the Q of the pendulum is not at least several hundred, then there is probably some interference from the connecting wires.

A geophone is a convenient choice for T in that it is both reciprocal and the inertial reaction of the driven proof mass is adequate to move the pendulum. However, it is not necessary that R (or S) be a geophone. If R is much less sensitive than the geophone T, though, a lock-in amplifier or some form of synchronous averaging may be necessary to maintain an adequate signal-to-noise ratio. If a two-channel spectrum analyzer is used, the coherence should be checked to verify a high signal-to-noise ratio. This is particularly important if random-noise excitation is being used. A drop in coherence indicates either a low signal-to-noise ratio or an overdriven sensor or electronics.

Two different fixtures were fabricated to demonstrate the technique. The first (fixture A) consisted of two pieces of aluminum channel to seat the geophones and form a symmetric structure. This fixture was 37 cm long, provided a geophone-to-geophone spacing of 18 cm, and had a mass of 1.38 kg including the geophones (but excluding the mass of the proof masses). Since this fixture held all three sensors, the entire calibration could be performed quickly.

Another fixture (fixture B) was made to accomodate a wider variety of sensors and to make the high-frequency performance more predictable. This fixture was a solid, cylindrical aluminum bar, 5 cm in diameter and 15 cm long, milled flat on both ends, and drilled and tapped axially at both ends for 10–32 threads. Holders with 10–32 studs were made for the geophones so that they could be easily interchanged. Although only two transducers at a time can be mounted, this arrangement permitted modeling of the bar as a waveguide at higher frequencies while still keeping the force-application axis coincident with the long axis of symmetry of the bar (to minimize excitation of bending modes). The mass of fixture B with two geophones in place was 1.042 kg.

Geospace GSC20-DH geophones[11] were used as S and T for horizontal-axis operation. These units have a 10-Hz resonance, an 8.9-g proof mass, an electrical resistance of 7600 Ω, a coil inductance of 900 mH, and an intrawinding capacitance of about 10 nF. As indicated from the calibrations below, the response above resonance was about 135 V/(m/s). As a consequence of Eq. (7), these geophones were also capable of producing 135 mN of reaction force per mA of drive current. The pendulum resonance was 0.65 Hz with a Q of 650. With the 1.36-kg fixture and 3-mA-rms of drive current (the maximum used in these measurements), the acceleration produced was 300 mm/s^2 or approximately 0.03 g. When selecting geophones for this calibrator, devices with a large α (usually a device specification) should be selected

not only to produce large received signals but also to produce large forces on the pendulum.

Fixture B was also used for calibrating vertical-axis sensors. Geospace GS30CT geophones were used and the cylinder was suspended vertically using a loop of surgical tubing. The geophones have a 10-Hz resonance, an 11.2-g proof mass, an electrical resistance of 2565 Ω, and a response above resonance of 70 V/(m/s). The pendulum resonance was 0.75 Hz with a Q of 18. Since fixture B only permits mounting two sensors at one time, the pendulum mass can be different for different measurements as allowed by Eqs. (12)–(14).

III. QUALITY CHECKS

There are at least four correctable errors: (1) electromagnetic coupling from the driving geophone to the receiving geophone must be reduced to a level well below that of mechanical coupling through the pendulum; (2) capacitive coupling between the driver and the receiver through the cases and through the pendulum must be minimized by careful grounding; (3) the dynamic behavior of the pendulum must be known well enough that corrections near the pendulum resonance or wave corrections at high frequency can be made and spurious resonances must be minimized; and (4) the true open-circuit voltage must be determined either by buffering the sensor output or by calculating corrections for loading. Once these errors have been identified and minimized, the apparatus can be directly tested for the degree to which it is reciprocal over the band of interest.

Electromagnetic coupling should be checked by firmly clamping the pendulum bar (or detaching the geophones and placing them on *separate* soft foam pads at the same spacing as they would be on the bar) and performing the measurement. The received signals should be well below those obtained with the bar free. For several varieties of geophones, a separation of 15 cm was more than adequate to reduce the electromagnetic coupling to an acceptable level. From measurements with a coil (1200 turns of 32-gauge wire on a 3-cm-diam form), the radiated field from the geophone (for coil and geophone coaxial) was seen to decay as center-to-center distance to the (minus) fourth power: small increases in spacing produce large decreases in coupling. Replacing the geophones by equivalent resistors should also be done to identify any crosstalk paths other than sensor to sensor.

In the arrangement suggested, capacitive coupling between the source and receive geophones normally appears as a difference in the output voltage when the connection to the receive geophone is reversed. (The capacitive coupling adds to the real signal in one connection and subtracts from the real signal in the other connection.) This coupling can be reduced to acceptable levels by running a separate ground connection from the electronics to the pendulum body. Once electromagnetic and capacitive coupling have been minimized, it is useful to perform a test measurement and calculate the coherence between drive current and received voltage at a number of frequencies. For subpercent accuracy, the coherence should be maintained above 0.999.

Random-noise excitation of the source geophone is useful as a quick check for spurious behavior in the apparatus. If the frequencies of the resonances of the geophones and test accelerometer are roughly known, other features in the received spectrum may be spurious resonances in the pendulum bar assembly. If the three sensors are clamped between two bars (as in fixture A), this test is important since structural resonances below 1 kHz are possible. If in-band resonances cannot be eliminated by redesign of the bars, an apparatus similar to fixture B should be used.

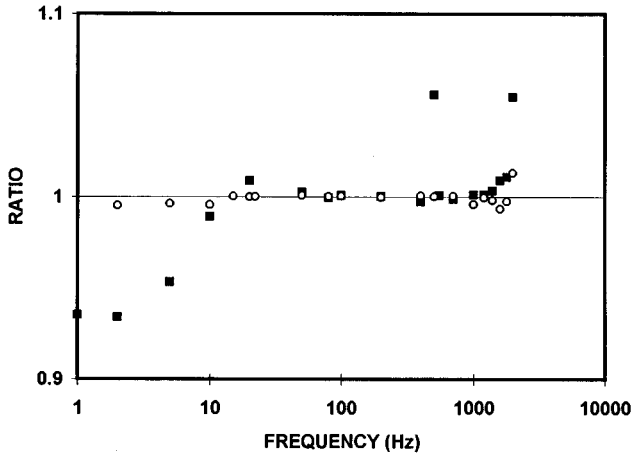


FIG. 5. Reciprocity tests for: (a) fixture A oriented horizontally (black squares), and (b) fixture B oriented vertically (open circles).

Since two geophones are used in this technique, the reciprocal nature of the apparatus should be checked directly. This is, perhaps, the most important test of the apparatus. At discrete frequencies throughout the band of interest, the ratio of received voltage from T to drive current into S should be measured (along with the coherence between the two channels). The roles of the two geophones should then be reversed and, at the same frequencies, the ratio of the received voltage from S to drive current into T should be measured. These two ratios should, ideally, be identical and, in practice, are generally within a few tenths of a percent. Departures from reciprocal behavior are indicated by larger differences. For example, Fig. 5 shows a comparison between such measurements using fixture A (using GSC20DH geophones) and fixture B (using GS30CT geophones). For fixture A, there is a clear departure from reciprocity below 20 Hz. This was probably the result of nonlinearity in the GSC20DH geophones and asymmetrically excited resonances in the fixture.

IV. PENDULUM AND GEOPHONE DYNAMICS

The assumption that the pendulum acts solely as a free mass is not quite correct. Even at frequencies for which the pendulum can be treated as a lumped element, the dynamics of the geophones that are attached affect the transfer impedance Z_P. An equivalent circuit[12] for the pendulum with one geophone attached is shown in Fig. 3, using the mobility analogy for the mechanical side. The electrical components are included but do not influence the solution for Z_P when the electrical terminals are open-circuited and the frequency is low enough so that the intrawinding capacitance C_e acts as

an open circuit. The geophone case mass is included in the pendulum mass m_P; the mass m_T refers to the geophone proof mass only.

From the circuit in Fig. 3, the transfer impedance can be written as follows:

$$Z_P = j\omega m_P \left[\left(1 - \frac{\omega_p^2}{\omega^2} - j\,\frac{\omega_P}{\omega Q_P} \right) + \frac{m_T}{m_P} \left(\frac{1 + j\omega/\omega_T Q_T}{1 - \omega^2/\omega_T^2 + j\omega/\omega_T Q_T} \right) \right], \tag{16}$$

where ω_T and ω_P are the resonance frequencies of the geophone and the pendulum, respectively, and Q_T and Q_P are the corresponding Q's of those resonances. Since it is easy to achieve a pendulum resonance below 1 Hz (a 1-m-long pendulum oscillates at about 0.5 Hz) with a Q of several hundred, the first factor in parentheses represents a correction of less than 1% for the receiving response above a few hertz and less than 0.1% above 20 Hz. By using Eq. (16) instead of Eq. (15), good precision can be obtained at frequencies as low as twice the pendulum frequency.

The second factor in parentheses in Eq. (16) starts at unity well below the geophone resonance (typically 10 to 30 Hz), peaks near the geophone resonance, and drops linearly with frequency above the resonance. The Q of a geophone is normally slightly less than one, which means that the maximum magnitude of this correction is less than 1.5. The relative influence of this second correction is determined primarily by the ratio of geophone proof mass to pendulum mass. With a mass ratio, m_P/m_T, of 50:1, the maximum error resulting from omitting this term is less than 3%. (Note: m_T is the total mass of the proof masses for *all* of the geophones attached to the pendulum.) It should be kept in mind that raising the pendulum mass decreases the signal at the receiving transducer so the mass ratio should not be made too high. It is better to include the mass correction and maintain a high signal-to-noise ratio than to reduce excessively the amplitude of the mass motion by increasing the pendulum mass.

At low frequencies where the geophone dynamics become important, use of Eq. (16) for the transfer impedance should be considered. It might be argued that this technique then requires accurate knowledge of many physical parameters; however, the sensitivity of Eq. (16) to any parameter other than the pendulum mass is low. For example, the proof mass is easily known to within 10% and a 10% error in that quantity produces a *maximum* error in the transfer impedance (for fixture B) of 0.6% at resonance and only 0.08% an octave higher.

If two geophones are attached to the apparatus, Eq. (16) can still be used by assuming that both geophones have identical dynamics (doubling m_T, for example). The exact solution has been computed for several cases using individual geophone dynamics; however, there have been no situations in which the errors introduced by the assumption of identical dynamics are comparable to the basic measurement errors. Further refinement of the geophone dynamics may be possible by making measurements with different pendulum masses,[3] but the improvement in results may not justify the effort. Accuracy of 1% or better can be obtained over much of the useful frequency range of the pendulum even with Eq. (15).

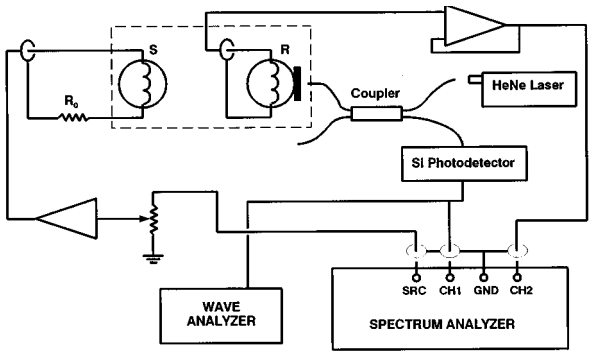


FIG. 6. Schematic of electrical connections for the Bessel-null calibration. The transducers S and R are rigidly attached to the pendulum (not shown). The amplifier shown on the SRC output is optional and can also be used to advantage in the reciprocity calibration.

An estimate of the frequency at which the pendulum bar can no longer be treated as a lumped mass can be made from simple waveguide theory.[13] By translating a zero-mass (dummy) load from one end of a short bar to the other, the effective mass of the bar including the waveguide correction can be found. The expression for small kx is

$$m_{\text{eff}} \approx \rho V \frac{\tan(kx)}{kx}, \tag{17}$$

where ρ is the density of the bar, V is the volume of the bar, k is the wave number in the bar (ω/c), and x is the length of the bar. The factor ρV is the actual mass. The effective mass is 1% different (greater) than the actual mass for $kx = 0.17$. For a sound speed of 5050 m/s (aluminum) and a length of 15 cm (fixture B), this would correspond to a frequency of 900 Hz. The first-order wave correction is easily incorporated into Z_p if desired by applying the $\tan(kx)/(kx)$ factor to Eq. (15) or (16).

V. COMPARISON WITH INTERFEROMETRIC CALIBRATION

In order to explore the ultimate accuracy of the reciprocity technique, a Bessel-null optical interferometric technique was used to check the reciprocity results. The Bessel-null method is particularly suited to such a comparison since it does not require maintaining a particular static path difference and is insensitive to drifts in the absolute value of the photodetector output. Consequently, the apparatus can be set up to operate on a free pendulum and the optical calibration can be done with the same experimental setup as used for reciprocity.

The optical arrangement is a standard fiber-optic Fabry–Perot interferometer (see Figs. 6 and 7) with a helium–neon laser (632.8 nm) and a silicon photodetector. The signal of interest is provided by interference between the beam reflected from the cleaved end of the sense fiber (and staying inside the fiber) and the beam exiting the sense fiber, bouncing off the mirror attached to the pendulum, and re-entering the fiber. The calibration procedure is essentially the stan-

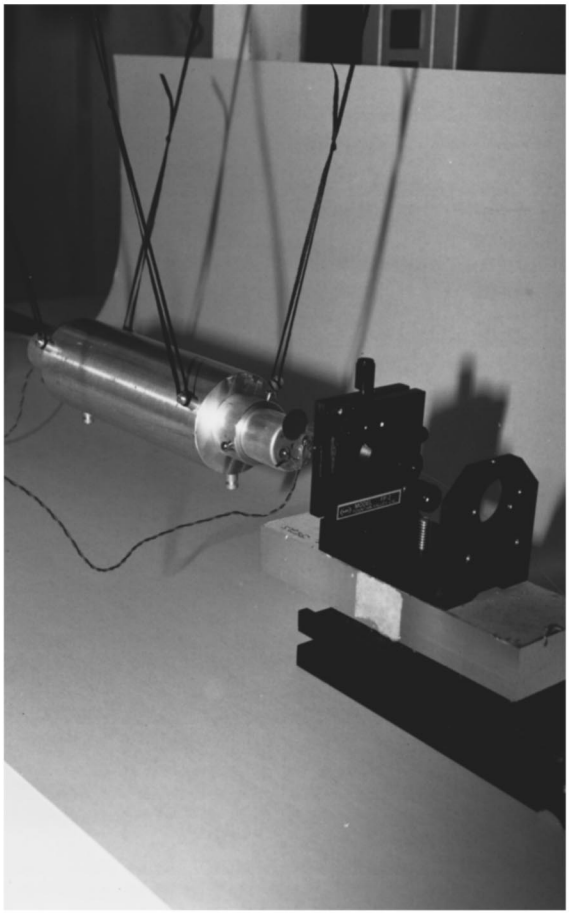


FIG. 7. Fixture B (cylinder with two geophones) in horizontal orientation with four-point suspension. Also shown on the right is the fiber fixture for performing the Bessel-null calibration.

dard Bessel-null technique[9] with the exception that the free oscillations of the pendulum provide a dither signal that makes controlling the static displacement unnecessary.

The frequency range over which the Bessel-null method is, in this case, applicable is limited on the low end (to about 20 Hz) by the excitation of the pendulum by external vibration (noise). The frequency range is limited on the high end (to about 140 Hz) by the ability of the driver to produce sufficient displacement amplitude in the pendulum to reach at least the first null. For example, in fixture A at 50 Hz, the first 15 nulls were obtained with a drive level from 0 to 25 V rms (0 to 3 mA) whereas, at 100 Hz, only six were obtained and, at 200 Hz, none could be obtained. Some improvement in frequency range is possible by isolating the system from external vibrations (to reduce the lower limit) and by using a lighter pendulum or more powerful drivers (to extend the upper limit).

The results of a typical comparison between the Bessel-null method and the reciprocity method (where a geophone is being calibrated) are shown in Table I and Fig. 8. The stability of the calibration value (at 100 Hz) was monitored over several days and amounted to a variation of ± 0.0004 for reciprocity and ± 0.0005 for the Bessel-null method (see Fig. 9).

In general, the desire to measure as many nulls as possible at a given frequency and the sensitivity of the Bessel-null method to vibrational noise make that technique considerably more time consuming than the reciprocity method and the reciprocity technique can be performed over a wider range of frequency. While the uncertainty estimated in the next section is higher for the Bessel-null method than it is for reciprocity, at a frequency low enough to have many obtainable nulls but still well above the low-frequency vibrational noise, excellent consistency and repeatability can be obtained optically. For example, at 80 Hz, nine nulls were reached with a standard deviation of only 0.03%.

TABLE I. Comparison of the response of a geophone measured by reciprocity and by Bessel null. (Error is given as $\pm 2\sigma$.)

Frequency (Hz)	Reciprocity [V/(m/s)]	Bessel null [V/(m/s)]
50	140.08 ± 0.14	140.47 ± 1.41
70	138.78 ± 0.14	138.98 ± 1.25
80	138.34 ± 0.14	138.61 ± 1.11
90	138.08 ± 0.14	138.34 ± 1.11
100	137.93 ± 0.14	138.11 ± 1.10
110	137.76 ± 0.14	137.98 ± 1.10
130	137.92 ± 0.14	137.71 ± 1.24
150	137.94 ± 0.14	137.74 ± 1.38

VI. SUMMARY OF SOURCES OF ERROR

The following section is a superficial examination of sources of error in the two techniques. Sufficient statistical information was not obtained for several of the error sources so this should not be considered rigorous. The analysis is sufficient to indicate the potential of the two techniques and to illustrate some of the areas of sensitivity.

A. Reciprocity

If the reciprocal transducer is not truly reciprocal, an error will be introduced into the calibration. The test for reciprocity is described in Sec. III (see also Fig. 5). The ratio of the current-to-voltage transfer functions in both directions should be one. Above 10 Hz and below 1 kHz, the departure from unity was typically less than ± 0.0002; the response is proportional to the square root of the transfer function so the

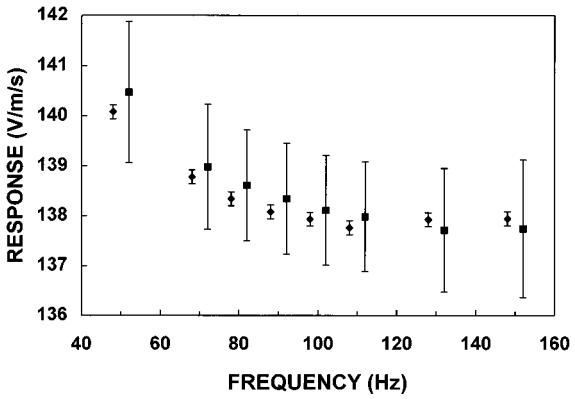


FIG. 8. Comparison of reciprocity (diamonds) and Bessel-null (squares) calibration at several frequencies. Each set is offset around the relevant frequency for clarity. Error bars indicate absolute uncertainty.

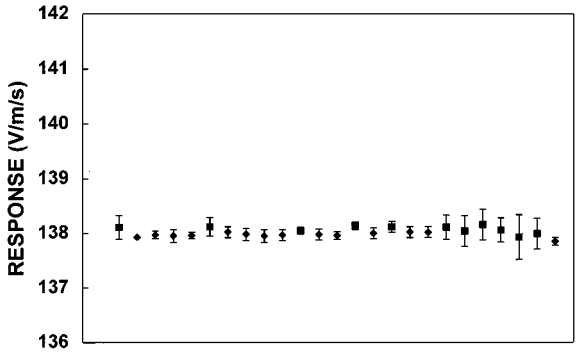


FIG. 9. Repetitive measurement at 100 Hz over a 1-week period by reciprocity (diamonds) and Bessel-null (squares). Error bars show the variation in measurement only, not absolute uncertainty. The larger variation in the last several Bessel-null values resulted from taking only one measurement per null point.

effective error is ± 0.0001 over most of the band of interest. No values were accepted if the reciprocity departure was greater than ± 0.002 and such large departures were only considered for measurements below 20 Hz. Consequently, the lowest frequency data presented here can have effective reciprocity-induced errors as high as ± 0.001.

The two obvious sources of error in the reciprocity technique are errors in measuring the mass of the pendulum apparatus and errors in measuring the reference resistor by which the current into the driver is measured. The mass can be measured easily to within 1 g for the 1.36-kg apparatus or ± 0.00075. Since the response is proportional to the square root of the mass, the effective error is ± 0.0004. The resistance, measured in circuit with the four-wire HP3468A multimeter, is determined to within 0.01 Ω out of the 100-Ω value or ± 0.0001. The temperature coefficient of the resistor was 100 ppm/ °C and the temperature of the apparatus did not vary by more than ± 2 °C throughout the measurements, which would contribute a variation of ± 0.0002. (The maximum power dissipated in the 0.25-W resistor was 0.0006 W.) As in the case of the mass, the response is proportional to the square root of the resistance; therefore the effective error is ± 0.0001.

Because all of the measurements involved in the reciprocity calibration are voltage ratios, the absolute accuracy of the voltmeter or analyzer is not critical. For these measurements, a two-channel FFT analyzer (HP35660) was used with one channel to monitor the voltage across the reference resistor and one channel to monitor the voltage output of the receive transducer. In this case (where the same electronics are not used to measure both channels), the ratio error between the two channels contributes to the overall error budget. This error was maximum for the lowest voltage range used at $+0.0008$ for channel 2 relative to channel 1. Over most of the relevant voltage range, the ratio error was only 0.0001 ± 0.0001. Since the reciprocity values were always measured at 3–4 drive levels and averaged, the ratio error will be treated arbitrarily as a random error of ± 0.0004. After the square root, the effective error is ± 0.0002.

The measurement electronics can load the device being measured. This was corrected in the calculation of response instead of interposing high-impedance buffer amplifiers (which would, themselves, need to be calibrated). The parallel combination of oscilloscope and analyzer presents a 500 000-Ω input to either the 100-Ω reference resistor or the 7600-Ω geophone. The true voltage across the reference resistor is 1.0002 times the voltage measured and the true voltage across the geophone is 1.015 times the voltage measured. The actual resistances of the geophones were $\pm 70\ \Omega$ around the nominal 7600 Ω and this produced an error in the assumed loading correction of ± 0.00015. For frequencies greater than a few hundred hertz, the reactive component of the geophone source impedance should be considered in the loading calculation. A better strategy, though, would be to use a unity-gain, high-impedance buffer amplifier on the output of the receive geophone as shown schematically in Fig. 4.

For the frequency range used in the optical-to-reciprocity comparison (20–150 Hz), the error resulting from stray coupling between the source and the receiver was below 0.0002. This is a systematic deviation (tending to make the reciprocity results slightly high for normal-polarity connections) and for simplicity it will be ignored at frequencies below 100 Hz and taken to be increasing from 0 at 100 Hz to $+0.0002$ at 150 Hz. At 1 kHz, this coupling can be 0.005 for the simple beam apparatus although it can be reduced by more than an order of magnitude from the measurement by averaging the voltage obtained with normal connections to that obtained with reversed connections (at the receiving transducer).

Combining these effective errors (square root of the sum of the squares) and multiplying by two (for the 2σ point as if the errors were true standard deviations) produces an estimate for the uncertainty in the reciprocity results of $\pm 0.001(\pm 0.1\%)$ above 10 Hz. For lower frequencies, the reciprocity-departure error can dominate and bring the total (2σ) error up to $\pm 0.002(\pm 0.2\%)$.

B. Bessel null

Unlike the reciprocity technique, the Bessel-null method depends on an absolute measurement of the output voltage of the receiving transducer so the accuracy of the voltmeter or analyzer is an important component of the error budget. The HP35660 spectrum analyzer is only rated to within a few percent for absolute accuracy but with sufficient warm-up time, it is *stable* in response to a much tighter tolerance. The analyzer was checked against an HP3458A multimeter (± 0.003) and, over the voltage range used in the calibration, the 35 660 to 3458 comparison error ranged from -0.004 at the lowest voltage used to $+0.003$ at the high end of the voltage range. For most of the frequencies, a number of nulls (corresponding to different voltage drive levels) were used to estimate the response so the absolute error was reduced by averaging. For the upper frequencies, where only one or two nulls were used, the voltages tended to be in the middle of the range where the absolute error was low. Consequently, this error will be treated as a random error of ± 0.004 (including the error of the multimeter). This error dominates the overall error budget.

TABLE II. Reciprocity calibration of PCB393A31 piezoceramic accelerometer in vertical and horizontal configurations of the free-mass calibrator. Bessel-null calibration data and calibration data from NIST are included for comparison. (The uncertainty in the NIST data is given as 2% below 5 Hz and 1% from 5 to 160 Hz.)

Frequency (Hz)	Horizontal (volts per g)	Vertical (volts per g)	Bessel null (volts per g)	NIST (volts per g)
1	[a]	[b]	[c]	10.56
2	[a]	10.51	[c]	10.49
5	[a]	10.49	[c]	10.41
10	[a]	10.39	[c]	10.30
20	10.34	10.27	10.12	10.26[e]
50	10.27	10.22	10.14	10.19
70	10.24		10.17	
80	10.26		10.18	
90	10.26		10.21	
100	10.26	10.18	10.22	10.11
115	10.28		10.13	
130	10.29		10.26	
150	10.34	10.25	[d]	10.17[e]
200	10.46	10.39	[d]	N/A

[a]Non-negligible departure from reciprocity for the horizontal geophones.
[b]The drive level achieved with the vertical geophone was marginal at this frequency; the coherence was below 0.999.
[c]Insufficient signal-to-noise ratio at these frequencies for Bessel null.
[d]Insufficient drive amplitude to reach first null.
[e]NIST calibration was not performed at these frequencies; values obtained by interpolation.

There is additional uncertainty in determination of the null points. At the extremes of the frequency range for the optical calibration (50 and 150 Hz), this uncertainty amounted to ± 0.003 while at 100 Hz the uncertainty was ± 0.001. These uncertainties can be at least halved by making multiple determinations of each null and by using only higher-order nulls since the signal is further above the background noise for each successively higher null.

Because a helium–neon laser was used for the optical calibration, the uncertainty in the wavelength (about 0.000002) is negligible; however, for a semiconductor laser, this source of uncertainty might be significant. Also, the source-to-receiver stray coupling is negligible in comparison to the absolute voltage measurement error and the null-point error. Combining these effective errors and multiplying by two gives an overall uncertainty (dominated by the error in absolute voltage measurement) of $\pm 0.01(1\%)$ at 50 and 150 Hz reducing to $\pm 0.008(0.8\%)$ in between. Below 50 Hz, obtaining an adequate signal-to-noise ratio becomes increasingly difficult in the Bessel-null method (on a free pendulum); the uncertainty at 20 Hz is at least $\pm 0.025(2.5\%)$.

VII. CALIBRATION OF A PIEZOCERAMIC ACCELEROMETER

In order to examine the performance of the calibrator at lower frequencies (in particular, through the resonance of the geophones), a commercial piezoelectric accelerometer (PCB 393A31) was calibrated by the reciprocity technique described above. Fixture A was used in the horizontal orientations and fixture B was used in the vertical orientation. The impedance correction of Eq. (21) was applied below 50 Hz. The results are summarized in Table II and in Fig. 10 along with results obtained by Bessel-null (horizontal axis) and a recent NIST calibration[14] (vertical axis).

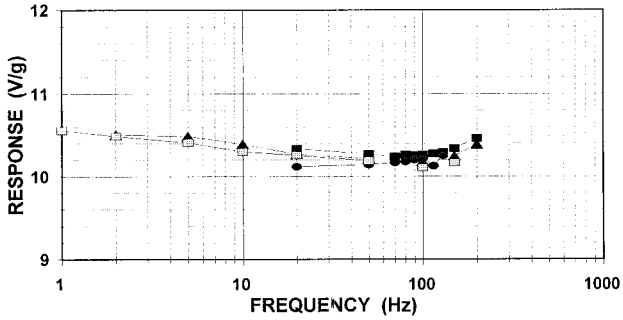


FIG. 10. Calibration of piezoelectric accelerometer (PCB 393A31). Results are given for vertical-axis reciprocity (triangles), horizontal-axis reciprocity (dark squares), Bessel-null (circles), and NIST (light squares).

The vertical-axis reciprocity results and the vertical-axis NIST results are well within the quoted uncertainty of the NIST results (1% at 5 Hz and above, 2% below 5 Hz) in spite of the fact that the reciprocity calibration was performed above, at, and below the geophone resonance. The horizontal-axis reciprocity results are systematically higher than the vertical values (by 0.7% on average). This departure is greater than the uncertainty of the calibration. The accelerometer may have a small orientational dependence in response but further tests should be run before accepting the difference as real. For example, if the uncertainty of the Bessel-null method were reduced through the use of a more nearly accurate voltage measurement, the optical technique could be applied to both the horizontal and vertical configurations. As implemented here, the uncertainty in the Bessel-null method is too high to resolve unambiguously a 0.7% difference.

Since 1% accuracy is acceptable for many calibrations (1% error is equivalent to 0.09 dB), the free-mass reciprocity technique is clearly valuable for low-frequency calibration of vibration sensors. Because of the good signal-to-noise ratios obtainable with the geophones, the technique is particularly valuable for small-signal calibrations. Because of the simplicity and rigidity of the apparatus, it is also valuable for horizontal-axis calibration of heavy sensors.

ACKNOWLEDGMENTS

Steven L. Garrett (The Pennsylvania State University) suggested the use of the optical Bessel-null technique as a check on the reciprocity method. Howard Krumholtz (NAWC Aircraft Division, Warminster) set up the laser, fiber coupler, and photodetector for the Bessel-null calibration and made a number of comparison measurements between small-signal interferometry and the Bessel-null method.

[1] I. Rudnick, ''Unconventional reciprocity calibration of transducers,'' J. Acoust. Soc. Am. **63**, 1923–1925 (1978).
[2] G. W. Swift and S. L. Garrett, ''Resonance reciprocity calibration of an ultracompliant transducer,'' J. Acoust. Soc. Am. **81**, 1619–1623 (1987).
[3] S. Levy and R. R. Bouche, ''Calibration of vibration pickups by the reciprocity method,'' J. Res. Natl. Bur. Stand. **57**, 227–243 (1956).
[4] M. Harrison, A. O. Sykes, and P. G. Marcotte, ''The reciprocity calibra-

tion of piezoelectric accelerometers,'' J. Acoust. Soc. Am. **24**, 384–389 (1952).
[5] A. London, ''The absolute calibration of vibration pickups,'' NBS Tech. News Bull. **32**, 8–10 (1948).
[6] H. M. Trent, ''The absolute calibration of electromechanical pickups,'' J. Appl. Mech. **15**, 49–52 (1948).
[7] C. T. Morrow, ''Reciprocity calibration of vibration probes,'' J. Acoust. Soc. Am. **20**, 826–829 (1948).
[8] S. P. Thompson, ''Reciprocity calibration of primary vibration standards,'' J. Acoust. Soc. Am. **20**, 637–640 (1948).
[9] R. Bouche, *Calibration of Shock and Vibration Measuring Transducers* (The Shock and Vibration Information Center, Naval Research Laboratory, Washington, DC, 1979), Shock and Vibration Monograph 11.
[10] T. ten Wolde, ''On the validity and application of reciprocity in acoustical, mechano-acoustical and other dynamical systems,'' Acustica **28**, 23–32 (1973).
[11] In this paper, a number of commercial products are mentioned. This should not be construed as an endorsement of those particular products by the U.S. Government or any of its agencies. In most cases, satisfactory products are available from other manufacturers.
[12] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982), 3rd ed., Secs. 14.3 and 14.4.
[13] W. H. Hayt, *Engineering Electromagnetics* (McGraw-Hill, New York, 1967), 2nd ed. Sec. 12.1; use the mechanical impedance analogy to convert electrical quantities to mechanical quantities.
[14] National Institute of Standards and Technology, 28 November 1995. This is a standard NIST calibration product and does not represent the lowest uncertainty available through custom services.

Space-time array processing: The model-based approach

E. J. Sullivan
Code 82101, Naval Undersea Warfare Center, Newport, Rhode Island 02841

J. V. Candy
University of California, Lawrence Livermore National Laboratory, P.O. Box 808, L-156, Livermore, California 94550

(Received 10 September 1995; accepted for publication 5 August 1997)

A method of space–time array processing is introduced that is based on the model-based approach. The signal and measurement systems are placed into state-space form, thereby allowing the unknown parameters of the model, such as signal bearings, to be estimated by an extended Kalman filter. A major advantage of the model-based approach is that there is no inherent limitation to the degree of sophistication of the models used, and therefore it can deal with other than plane-wave models, such as cylindrically or spherically spreading propagation models, as well as more sophisticated representations such as the normal mode and the parabolic equation propagation models. Since the processor treats the parameters of interest as unknown parameters to be estimated, there is no explicit beamformer structure, and therefore no accuracy limitations such as fixed beam bin sizes and predetermined number of preformed beams. After a theoretical exposition of the underlying theory, the performance of the processor is evaluated with synthesized data sets. The results indicate that the method is a highly effective approach that is capable of significantly outperforming conventional array processors.
[S0001-4966(97)06111-0]

PACS numbers: 43.60.Gk, 43.30.Wi [JLK]

INTRODUCTION

Underwater acoustic array processing requires that one must assume some form for the signal and noise models. In the case of bearing estimation, a plane wave is usually chosen for the signal model. For example, the conventional beamformer embodies an implicit assumption of such a model in the so-called steering or wavefront vector. Our goal in this paper is to recognize the reliance of array processing on such models and to exploit them for the purpose of fully utilizing the inherent information contained in them.[1]

Although a plane-wave model is used in this study, it will become clear that the model-based formalism permits the use of much more highly sophisticated models, such as the normal mode and parabolic equation propagation models.[2] The concept of using models in signal processing schemes is not a new idea, but probably originated with Hinich,[3] who showed that the depth of an acoustic source could be inferred from the measurements of a vertical array in conjunction with a propagation model. This led to what is now known as matched-field processing.[4,5] Recently, this concept has been formalized, leading to what we refer to here as model-based processing (MBP). Recent work in MBP has applied it to several problems in ocean acoustic signal processing. For more information, see Ref. 6 and references therein. The results of these studies have shown MBP to be a very effective technique, since it allows the introduction of the signal and propagation models directly into the signal processing structure in a self-consistent manner. The approach is based on casting the problem into state-space form. There are basically two sets of equations in the MBP representation: the state equations themselves, and the measurement equations, which relate the states to the actual measured quantities. In the stochastic case, a Gauss–Markov representation evolves, which allows the inclusion of both the measurement noise and the state or ''process'' noise as second-order statistical models. Once this framework is established, it is possible to investigate ocean acoustic signal processing problems in a very general way, since full advantage can be taken of existing systems theory (e.g., observability, identifiability) and techniques (Kalman state estimators) to characterize a recursive processor which constitutes a minimum-variance, adaptive estimator.[7] That is, the so-called innovation sequence, which is automatically generated by the Kalman filter algorithm, allows the performance of the model to be monitored, while at the same time Kalman estimators can be utilized on-line to continuously update model parameters, such as the sound speed profile, thus making inroads into the so-called *mismatch problem*.[8]

In this paper, we consider a simple plane-wave propagation model and apply it to the problem of bearing estimation. Although, as mentioned above, we are clearly not limited to this particular signal model, we feel that for this problem it is sufficiently realistic, and therefore enables the power of the MBP formalism to manifest itself without being shrouded by an unnecessarily complicated model. In particular, we will see that this approach casts the bearing estimation problem into a rather general form that *eliminates* the need to introduce an explicit beamformer structure, while evolving into a passive synthetic aperture (PASA)[9–12] structure in a natural way. Although we are not interested in PASA *per se*, but in model-based array processing in general, it will become apparent that PASA forms a natural framework in which to view this approach, since it is a time-evolving spatial process involving a moving array. The issue of the motion is an

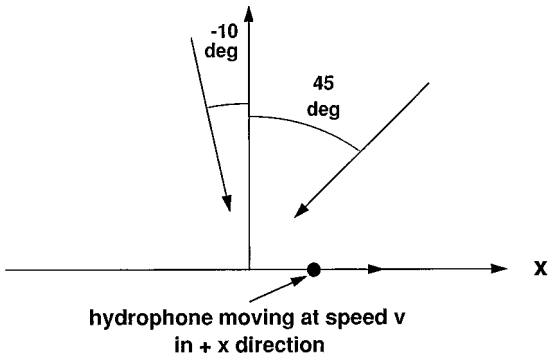


FIG. 1. Array and signal configurations for the single hydrophone case.

important one, since, as shown by Edelson,[13] the Cramer–Rao lower bound (CRLB) on the bearing estimate for a moving line array is less than that for the same physical array when not moving. In particular, the ratio of the CRLB for the moving array to that for the stationary array, which we denote by R, is given by

$$R=\mathrm{CRLB}_{\mathrm{moving}}/\mathrm{CRLB}_{\mathrm{fixed}}=[1+(3/2)(D/L)+(D/L)^2]^{-1}. \tag{1}$$

Here, D is the ''dynamic aperture,'' i.e., the speed of motion of the array times the total time, and L is the length of the physical aperture. As can be seen, there is the potential for a highly significant improvement in performance since R dramatically decreases as D increases.

In retrospect, the term ''synthetic aperture'' should probably be considered somewhat dated, since it tends to focus on the spatial aspects of array processing, whereas it is, in fact, a space–time process wherein the spatial and temporal aspects (such as ''array gain'' and ''signal-to-noise ratio'') are intimately entwined and cannot be separated from each other.

To motivate this approach, we begin with a single hydrophone receiver moving with speed v in the $+x$ direction, and a plane-wave signal model (see Fig. 1). The unit amplitude signal at the hydrophone, assuming that at $t=0$ the hydrophone is at $x=x_0$, is then given by

$$p(x,t)=e^{i\omega_0 t+i\beta(t)\sin\theta}, \tag{2}$$

where

$$\beta(t)=k_0(x_0+vt), \tag{3}$$

and $k_0=2\pi/\lambda$.

Suppose we now sample the signal in time. If we wish to estimate the signal bearing θ, one approach would be to assume a best guess for the initial value of θ and then recursively correct it by comparing the *measured* signal, $p(t_k)$, at each sample time t_k to its value at t_{k-1}, i.e., to assume that

$$\hat{\theta}(t_k|t_k)=\hat{\theta}(t_k|t_{k-1})+\Delta\theta(t_k). \tag{4}$$

Here, the hats on $\hat{\theta}(t_k|t_k)$ and $\hat{\theta}(t_k|t_{k-1})$ indicate that they are estimates. It is to be noted that we have generalized the notation somewhat in order to emphasize the recursive nature of the process. That is, $\hat{\theta}(t_k|t_{k-1})$ is the estimated (predicted) value of $\theta(t)$ at time t_k based on the data up to time t_{k-1}, and $\hat{\theta}(t_k|t_k)$ is the corrected value of $\theta(t)$ where the correction is based on the *measured* value of the pressure, $p(t_k)$ with the measurement being taken at $t=t_k$. Similarly, $\hat{p}(t_k|t_{k-1})$ is the predicted value of the pressure based on the data up to time t_{k-1}, i.e.,

$$\hat{p}(t_k|t_{k-1})=e^{i\omega_0 t_k+i\beta(t_k)\sin\hat{\theta}(t_k|t_{k-1})}. \tag{5}$$

The problem, then, is how to determine the correction $\Delta\theta(t_k)$ in some self-consistent manner. Since the pressure is function of θ and is parameterized by t_k, a first-order Taylor expansion of the pressure field at the hydrophone about $\theta=\hat{\theta}(t_k|t_{k-1})$ results in the following approximation for the measurement:

$$p(\theta)\approx\hat{p}(\hat{\theta}(t_k|t_{k-1}))+\left\{\left.\frac{\partial p(\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}(t_k|t_{k-1})}\right\}\Delta\theta(t_k), \tag{6}$$

where we have suppressed the dependence on t_k for clarity and now explicitly show the dependence of the pressure on θ. Solving for $\Delta\theta(t_k)$ yields

$$\Delta\theta(t_k)=\left\{\left.\frac{\partial p(\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}(t_k|t_{k-1})}\right\}^{-1}[p(\theta)-\hat{p}(\hat{\theta}(t_k|t_{k-1}))]. \tag{7}$$

We now define

$$K_\theta:=\left\{\left.\frac{\partial p(\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}(t_k|t_{k-1})}\right\}^{-1}. \tag{8}$$

Using Eq. (5) to eliminate $p(\theta)$, we have

$$K_\theta=\frac{-i\beta(t_k)\cos\hat{\theta}(t_k|t_{k-1})}{\beta^2(t_k)\cos^2\hat{\theta}(t_k|t_{k-1})}\,e^{-i\omega_0 t_k-i\beta(t_k)\sin\hat{\theta}(t_k|t_{k-1})}. \tag{9}$$

Note that $K_\theta=K_\theta(t_k)$, i.e., it is time varying. If, on the other hand, we choose the model-based approach to determine $\hat{\theta}$, then the resulting algorithm is a Kalman filter. In this case, returning to our original notation, we find that

$$\hat{\theta}(t_k|t_k)=\hat{\theta}(t_k|t_{k-1})+K(t_k)[p(t_k)-\hat{p}(t_k|t_{k-1})], \tag{10}$$

where the Kalman filter algorithm tells us that[7]

$$K(t_k)=\frac{-i\beta(t_k)\cos\hat{\theta}(t_k|t_{k-1})}{\beta(t_k)^2\cos^2\hat{\theta}(t_k|t_{k-1})+\sigma^2/\tilde{P}(t_k|t_{k-1})}\times e^{-i\omega_0 t_k-i\beta(t_k)\sin\hat{\theta}(t_k|t_{k-1})}. \tag{11}$$

Here, σ^2 is the measurement noise variance, i.e., the noise variance that would enter into the definition of the input or hydrophone level SNR, and $\tilde{P}(t_k|t_{k-1})$ is the predicted estimation error covariance, which is generated by the Kalman filter algorithm. Upon comparison of Eq. (11) with Eq. (9), we see that the first-order Taylor series approach leads to the Kalman gain for the case of vanishing measurement noise. In the case of more than one hydrophone, the generalization of Eq. (10) is

$$\begin{aligned}\hat{\theta}(t_k|t_k)&=\hat{\theta}(t_k|t_{k-1})+\mathbf{K}^{\mathbf{T}}(t_k)[\mathbf{p}(t_k)-\hat{\mathbf{p}}(t_k|t_{k-1})]\\&=\hat{\theta}(t_k|t_{k-1})+\mathbf{K}^{\mathbf{T}}(t_k)\boldsymbol{\epsilon}(t_k).\end{aligned} \tag{12}$$

$\mathbf{K}(t_k)$ is an $L\times 1$ vector, where L is the number of hydrophones, and $\hat{\mathbf{p}}(t_k|t_{k-1})$ is an $L\times 1$ vector of *predicted* measurements based on the data up to time$=t_{k-1}$. This equation tells us that the ''correction'' stage of the Kalman filter algorithm is based on the difference between the output of a beamformer with adaptive weight vector $\mathbf{K}(t_k)$ whose input is $\mathbf{p}(t_k)$, i.e., the actual pressure measurement vector taken at time$=t_k$, and the output of the same beamformer for input $\hat{\mathbf{p}}(t_k|t_{k-1})$, i.e., the ''predicted'' measurement. It is important to emphasize this point, since in the single hydrophone case our model-based bearing estimator has no inherent beamformer structure, but functions as a straightforward recursive parameter estimation procedure whereas the vector algorithm exploits the multiple hydrophone situation with a *naturally embedded* beamformer structure, i.e., as mentioned before, we impose no explicit beamformer structure in our formulation. Note also that for the multiple bearings case, Eq. (12) further generalizes to

$$\begin{aligned}\hat{\Theta}(t_k|t_k)&=\hat{\Theta}(t_k|t_{k-1})+\mathbf{K}^{\mathbf{T}}(t_k)[\mathbf{p}(t_k)-\hat{\mathbf{p}}(t_k|t_{k-1})]\\&=\hat{\Theta}(t_k|t_{k-1})+\mathbf{K}^{\mathbf{T}}(t_k)\epsilon(t_k).\end{aligned}\tag{13}$$

In this case, $\mathbf{K}(t_k)$ enlarges to an $L\times M$ matrix and $\hat{\Theta}(t_k|t_k)$ is an $M\times 1$ vector, with M being the number of bearings to be estimated.

This development offers a compelling argument for applying the model-based approach to array processing since, among other things, it allows the estimation of parameters of interest using well-established estimation methods while also providing access to all of the accompanying statistical theory that has been developed over a period of several years.

As mentioned above, one can choose to view this application of MBP as a form of passive synthetic aperture processing, since the motion of the array is explicitly contained in the signal model. As will be seen later, this results in a significant improvement in performance. It is important to emphasize this since in the past there was a conventional wisdom regarding PASA techniques, which held that practical limitations prevented them from being applicable to real-world systems. The issues are threefold: (1) since PASA can be viewed as a scheme that converts temporal gain to spatial gain, most signals of interest do not have sufficient temporal coherence to allow a long spatially coherent aperture to be synthesized; (2) since past algorithms required *a priori* knowledge of the source frequency in order to compute the phase correction factor, the method was essentially useless in any bearing estimation problem because Doppler would introduce an unknown bias on the frequency observed at the receiver; and (3) since synthetic aperture processing essentially converts temporal gain to spatial gain, there was no ''new'' gain to be achieved, and therefore, no point to the method. Recent work[10–12] has shown that all of these objections are either *not true* or involve *unrealistic* assumptions. (For a thorough discussion of these issues, see Ref. 12.) Here, then, we take the view that not only is PASA a viable method of array processing, but is in fact a natural way of viewing it, since it is a time-evolving spatial process, involving a moving array, that fits quite naturally into a MBP structure.

Finally we wish to point out that, when viewed as a synthetic aperture process, model-based array processing constitutes a *totally new* version of PASA, and not a generalization of previous methods.[12]

In the next section, we give an overview of the state-space formalism as applied to acoustic arrays. Section II covers the construction of the model-based processor. This is followed by some example problems using synthesized data in Sec. III while Sec. IV contains a general discussion of the limitations of bearing estimation in general. Finally, Sec. V contains some conclusions and suggestions for future work.

I. STATE-SPACE FOR ACOUSTIC ARRAYS

The solution to the wave equation as a function of space and time is any function of $t\pm x/c$, where t is time, x is position, and c is phase velocity along the x direction. For the narrow-band plane-wave case, a solution can be written in complex form as

$$f(x,t)=e^{i\omega(t\pm x/c)},\tag{14}$$

where ω is the radian frequency. Note that in the complex form, the solution is separable and therefore can be written as a *product* of a function of time only and a function of space only, i.e.,

$$f(x,t)=e^{i\omega t}e^{\pm i(\omega/c)x}=X(x)T(t).\tag{15}$$

The wave equation is given by

$$\frac{\partial^2 f(x,t)}{\partial x^2}-\frac{1}{c^2}\frac{\partial^2 f(x,t)}{\partial t^2}=0.\tag{16}$$

Substituting $X(x)T(t)$ for $f(x,t)$ in Eq. (16) results in

$$\frac{\partial^2 X(x)}{\partial x^2}\,T(t)-\frac{1}{c^2}\frac{\partial^2 T(t)}{\partial t^2}\,X(x)=0.\tag{17}$$

Dividing by XT (and ignoring the respective dependences of X and T on x and t for simplicity), we have

$$\frac{1}{X}\frac{d^2X}{dx^2}-\frac{1}{c^2}\frac{1}{T}\frac{d^2T}{dt^2}=0.\tag{18}$$

Following the standard separation of variables procedure we note that, since the first term on the left-hand side of Eq. (18) is a function of x only, and the second term is a function of t only, each term must equal a constant for Eq. (18) to be true. Thus

$$\frac{1}{X}\frac{d^2X}{dx^2}=-k^2\tag{19}$$

and

$$\frac{1}{c^2}\frac{1}{T}\frac{d^2T}{dt^2}=-\frac{\omega^2}{c^2},\tag{20}$$

where k and ω/c are constants. If we consider the situation to be a plane wave impinging on the x axis at an angle of θ degrees from broadside, then k becomes the ''trace'' wave number, i.e., the projection of $k_0=\omega/c_0$ on the x axis, where k_0 is the acoustic wave number and c_0 is the phase velocity in the medium. Thus $k=k_0\sin\theta$. The ''dispersion'' relation is

$$k^2=\frac{\omega^2}{c^2}. \tag{21}$$

The resulting ordinary differential equations are

$$\frac{d^2X}{dx^2}+k^2X=0, \tag{22}$$

and

$$\frac{d^2T}{dt^2}+\omega^2T=0. \tag{23}$$

Both equations are of the general form

$$\frac{d^2U}{du^2}+\alpha^2U=0. \tag{24}$$

Defining the state vector as

$$\mathbf{U}=\begin{bmatrix} U \\ \dot{U} \end{bmatrix},$$

the state space representation of Eq. (24) is

$$\frac{d}{du}\begin{bmatrix} U \\ \dot{U} \end{bmatrix}=\begin{bmatrix} 0 & 1 \\ -\alpha^2 & 0 \end{bmatrix}\begin{bmatrix} U \\ \dot{U} \end{bmatrix}. \tag{25}$$

Here, the dot signifies the derivative with respect to the independent variable u. Since we will find it convenient to treat the problem as one in real variables, we have the following state-space forms for Eqs. (22) and (23) with the R and I subscripts denoting the real and imaginary parts. The spatial equation [Eq. (22)] becomes

$$\frac{d}{dx}\begin{bmatrix} X_R \\ \dot{X}_R \\ X_I \\ \dot{X}_I \end{bmatrix}=\begin{bmatrix} 0 & 1 & 0 & 0 \\ -k^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -k^2 & 0 \end{bmatrix}\begin{bmatrix} X_R \\ \dot{X}_R \\ X_I \\ \dot{X}_I \end{bmatrix};$$

$$\begin{bmatrix} X_R(0) \\ \dot{X}_R(0) \\ X_I(0) \\ \dot{X}_I(0) \end{bmatrix}=\begin{bmatrix} 1 \\ 0 \\ 0 \\ -k \end{bmatrix}, \tag{26}$$

and the temporal equation [Eq. (23)] becomes

$$\frac{d}{dt}\begin{bmatrix} T_R \\ \dot{T}_R \\ T_I \\ \dot{T}_I \end{bmatrix}=\begin{bmatrix} 0 & 1 & 0 & 0 \\ -\omega^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\omega^2 & 0 \end{bmatrix}\begin{bmatrix} T_R \\ \dot{T}_R \\ T_I \\ \dot{T}_I \end{bmatrix};$$

$$\begin{bmatrix} T_R(0) \\ \dot{T}_R(0) \\ T_I(0) \\ \dot{T}_I(0) \end{bmatrix}=\begin{bmatrix} 1 \\ 0 \\ 0 \\ -\omega \end{bmatrix}. \tag{27}$$

The measurement equation is

$$p(x,t)=X(x)T(t), \tag{28}$$

where the real part is

$$p_R(x,t)=\mathrm{Re}[X(x)T(t)], \tag{29}$$

and the imaginary part is

$$p_I(x,t)=\mathrm{Im}[X(x)T(t)]. \tag{30}$$

If we fix x, then

$$p(x_j,t)=e^{ikx_j}[T_R(t)+iT_I(t)]. \tag{31}$$

If we fix t, then

$$p(x,t_k)=[X_R(x)+iX_I(x)]e^{i\omega t_k}. \tag{32}$$

Thus we see that if we model the states as X and T, then the measurement equation is nonlinear.

There are several possible approaches to the general acoustic array state-space problem leading to the following cases.

Case I. Use both state-spaces X and T.

$$\mathbf{p}(\mathbf{x},t_k)=\begin{bmatrix} p(x_1,t_k) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ p(x_L,t_k) \end{bmatrix}=\begin{bmatrix} \sum_m (X_R^m+iX_I^m)_1(T_R^m+iT_I^m) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \sum_m (X_R^m+iX_I^m)_L(T_R^m+iT_I^m) \end{bmatrix}. \tag{33}$$

Here, $\mathbf{x}=[x_1,\ldots,x_L]'$ and m labels the bearing/frequency pairs.

Case II. Use the spatial state space only and replace T with the model given by $e^{i\omega_m t_k}$, which is an eigensolution of Eq. (23):

$$\mathbf{p}(\mathbf{x},t_k)=\begin{bmatrix} p(x_1,t_k) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ p(x_L,t_k) \end{bmatrix}=\begin{bmatrix} \sum_m (X_R^m+iX_I^m)_1 e^{i\omega_m t_k} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \sum_m (X_R^m+iX_I^m)_L e^{i\omega_m t_k} \end{bmatrix}. \tag{34}$$

Case III. Use the temporal state space only and replace X with the model given by $e^{ik_m x_l}$, which is an eigensolution of Eq. (22) with $k_m x_l=k_0 x_l \sin\theta_m$:

$$\mathbf{p}(\mathbf{x},t_k)=\begin{bmatrix} p(x_1,t_k) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ p(x_L,t_k) \end{bmatrix}=\begin{bmatrix} \sum_m e^{ik_m x_1}(T_R^m+iT_I^m) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \sum_m e^{ik_m x_L}(T_R^m+iT_I^m) \end{bmatrix}. \tag{35}$$

For the case of a moving array, $k_m x_l$ would be replaced with $\beta_l \sin\theta_m$ [see Eq. (3)].

Case IV. Use the models given in Cases II and III for T and X, respectively, with a parameter estimation state space for the unknown bearing angles $\{\theta_m\}$. This is the form we shall use in this work, since it offers the simplest nontrivial approach to the bearing estimation problem. However, the development from Eq. (26) through Eq. (35) shows that the formalism can easily incorporate more complex dynamical models for the signal.

Defining the state vector θ as

$$\theta = \begin{bmatrix} \theta_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \theta_M \end{bmatrix}, \tag{36}$$

and the process noise vector as

$$\mathbf{w} = \begin{bmatrix} w_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ w_M \end{bmatrix}, \tag{37}$$

a Gauss–Markov model can be written as

$$\frac{d}{dt}\,\theta(t) = \mathbf{A}(t)\,\theta(t) + \mathbf{w}(t), \tag{38}$$

where $\mathbf{A}$ is the null matrix. The presence of the null matrix means that we are assigning no dynamics to the state vector θ, such as bearing rate or acceleration. If there were bearing dynamics, say bearing rate, present in the measurements, the Kalman filter would track it. The only reason one would include specific dynamics into the model would be if there were interest in directly estimating the relevant dynamic parameters. In fact, the model-based approach would allow a direct combination of bearing estimation and tracking by including the dynamic model of the source(s).

A discrete recursive Kalman filter structure now can be implemented for real-time application by discretizing Eq. (38) using a finite difference approximation. This results in

$$\frac{\theta(t_k) - \theta(t_{k-1})}{\Delta t} = \mathbf{w}(t_{k-1}), \tag{39}$$

which leads immediately to the following recursive estimator.

$$\hat{\theta}(t_k|t_{k-1}) = \hat{\theta}(t_{k-1}|t_{k-1}). \tag{40}$$

That is, the bearing angles are the states. Equation (40) constitutes the recursive form of the state equation for our array processor, where the associated measurement equation is given by

$$\mathbf{p}(\mathbf{x},t_k) = \begin{bmatrix} p(x_1,t_k) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ p(x_L,t_k) \end{bmatrix} = \begin{bmatrix} \sum_m e^{i\omega_m t_k + \beta_1(t_k)\sin\theta_m} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \sum_m e^{i\omega_m t_k + \beta_L(t_k)\sin\theta_m} \end{bmatrix}. \tag{41}$$

In a realistic problem, we would not only desire estimates of the bearings, but also the source frequencies and the signal amplitudes. These parameters will be included in the so-called *augmented* form of the Kalman filter which will be developed in the next section.

II. MODEL-BASED ACOUSTIC ARRAY PROCESSING

In this section we formulate the model-based solution to the space–time array processing problem by developing a general form of the model-based processor (MBP) design with various sets of unknown parameters. We define the *acoustic array space–time processing problem as:*

GIVEN a set of noisy pressure-field measurements and a horizontal array of L sensors; FIND the ''best'' (minimum error variance) estimate of source bearings, temporal frequencies, and amplitudes.

We choose to use the nonlinear extended Kalman filter (EKF) estimator in parameter estimation form[7,14] as our MBP. The general form of the algorithm for our problem is shown in Table I as a special case of the EKF. We use the nonlinear pressure-field measurement model, Eq. (41), which was developed previously for M monochromatic plane-wave sources. We will characterize each of the M plane-wave sources by a corresponding set of temporal frequencies, bearings, and amplitudes, $[\{\omega_m\},\{\theta_m\},\{a_m\}]$. That is,

$$p(x,t_k) = \sum_{m=1}^{M} a_m e^{i\omega_m t_k + \beta(t_k)\sin\theta_m}, \tag{42}$$

where x is the current spatial position along the x axis in meters, and v is the array speed (meters/sec). Here, $\beta(t_k) = k_0(x_0 + v\,t_k)$ as defined in Eq. (3).

If we further assume, as in Eq. (41), that the single sensor equation above is expanded to include an array of L sensors, then $x \rightarrow x_l$, $l = 1,...,L$ then we obtain

$$p(x_l,t_k) = \sum_{m=1}^{M} a_m e^{i\omega_m t_k + \beta_l(t_k)\sin\theta_m}. \tag{43}$$

This can be written in a more concise form as

$$\mathbf{p}(t_k) = \mathbf{c}(t_k,\boldsymbol{\Theta}), \tag{44}$$

where the vector $\boldsymbol{\Theta}$ represents the parameters of the plane-wave sources and $\mathbf{p}(t_k) = [p(x_1,t_k)p(x_2,t_k),\ldots,p(x_L,t_k)]^T$. The *augmented* Gauss–Markov state-space model then evolves as

TABLE I. Discrete extended Kalman filter parameter estimator.

Prediction	
$\hat{\Theta}(t_k\|t_{k-1})=\hat{\Theta}(t_{k-1}\|t_{k-1})$	(parameter prediction)
$\tilde{P}(t_k\|t_{k-1})=\tilde{P}(t_{k-1}\|t_{k-1})+R_{ww}(t_{k-1})$	(covariance prediction)
Innovation	
$\epsilon(t_k)=\mathbf{p}(t_k)-\hat{\mathbf{p}}(t_k)=\mathbf{p}(t_k)-\mathbf{c}[\hat{\Theta}(t_k\|t_{k-1})]$	(innovation)
$R_{\epsilon\epsilon}(t_k)=\frac{\partial}{\partial\Theta}\mathbf{c}[\hat{\Theta}(t_k\|t_{k-1})]\tilde{P}(t_k\|t_{k-1})\frac{\partial}{\partial\Theta}\mathbf{c}^T[\hat{\Theta}(t_k\|t_{k-1})]+R_{vv}(t_k)$	(innovation covariance)
Gain	
$\mathbf{K}(t_k)=\tilde{P}(t_k\|t_{k-1})\frac{\partial}{\partial\Theta}\mathbf{c}^T[\hat{\Theta}(t_k\|t_{k-1})]R_{\epsilon\epsilon}^{-1}(t_k)$	(Kalman gain or weight)
Correction	
$\Theta(t_k\|t_k)=\hat{\Theta}(t_k\|t_{k-1})+\mathbf{K}(t_k)\epsilon(t_k)$	(parameter correction)
$\tilde{P}(t_k\|t_k)=\left[I-\mathbf{K}(t_k)\frac{\partial}{\partial\Theta}c[\hat{\Theta}(t_k\|t_{k-1})]\right]\tilde{P}(t_k\|t_{k-1})$	(covariance correction)
Initial conditions	
$\hat{\Theta}(0\|0)\tilde{P}(0\|0)\frac{\partial}{\partial\Theta}\mathbf{c}[\hat{\Theta}(t_k\|t_{k-1})]:=\frac{\partial}{\partial\Theta}\mathbf{c}[\Theta]\Big\|_{\Theta=\tilde{\Theta}(t_k\|t_{k-1})}$	

$$\begin{bmatrix}\Theta(t_k)\\ \text{---}\\ \Omega(t_k)\\ \text{---}\\ A(t_k)\end{bmatrix}=\begin{bmatrix}\Theta(t_{k-1})\\ \text{---}\\ \Omega(t_{k-1})\\ \text{---}\\ A(t_{k-1})\end{bmatrix}+\mathbf{w}(t_{k-1}), \tag{45}$$

where $\Theta:=[\theta_1\cdots\theta_M]^T$, $\Omega:=[\omega_1\cdots\omega_M]^T$, $A:=[a_1\cdots a_M]^T$, and $\mathbf{w}$ is a zero mean Gaussian random vector with covariance R_{ww}. Thus the parameter vector for the multiple plane-wave model is

$$\mathbf{\Theta}:=\begin{bmatrix}\Theta\\ \text{---}\\ \Omega\\ \text{---}\\ A\end{bmatrix}, \tag{46}$$

for $\mathbf{\Theta}\in R^{3M}$, The following *augmented* state prediction equation evolves for our model-based processor then evolves as

$$\mathbf{\Theta}(t_k|t_{k-1})=\mathbf{\Theta}(t_{k-1}|t_{k-1}), \tag{47}$$

with Eq. (44) as the associated measurement equation.

Since the state-space model is linear with no explicit dynamics, the prediction relations are greatly simplified, while the correction equations become nonlinear due to the plane-wave measurement model of Eq. (44). This leads to the EKF wherein the nonlinearities are approximated with a first-order Taylor series expansion. Here we require the measurement Jacobian,

$$C(t_k,\mathbf{\Theta}):=\frac{\partial\mathbf{c}(t_k\mathbf{\Theta})}{\partial\mathbf{\Theta}}, \tag{48}$$

an $L\times 3M$ matrix. We simplify the notation by defining the following time varying coefficient,

$$\alpha_m(t_k):=a_m e^{i\omega_m t_k}. \tag{49}$$

The mth plane-wave source measured by the lth pressure-field sensor is now given by

$$p_l(t_k)=\sum_{m=1}^{M}\alpha_m(t_k)e^{i\beta_l(t_k)\sin\theta_m}. \tag{50}$$

Now we can calculate the required measurement Jacobian matrix of Eq. (44) from the relations for $l=1,...,M$ and

$$\begin{aligned}
&\frac{\partial\mathbf{c}(t_k,\mathbf{\Theta})}{\partial\theta_m}=i\alpha_m(t_k)\beta_l(t_k)\cos\theta_m e^{i\beta_l(t_k)\sin\theta_m},\\
&\qquad m=1,...,M\\
&\frac{\partial\mathbf{c}(t_k,\mathbf{\Theta})}{\partial\omega_m}=it_k\alpha_m(t_k)e^{i\beta_l(t_k)\sin\theta_m},\quad m=M+1,...,2M\\
&\frac{\partial\mathbf{c}(t_k,\mathbf{\Theta})}{\partial a_m}=e^{i(\omega_m t_k+\beta_l(t_k)\sin\theta_m)},\quad m=2M+1,...,3M.
\end{aligned} \tag{51}$$

Given Eqs. (44), (47), and (51), we are now in a position to compute a recursive (predictor/corrector form), extended Kalman filter estimate of the state vector $\mathbf{\Theta}$. Referring to Table I, the steps are as follows:

(1) Given an initial or trial value of the state estimate, $\hat{\Theta}(t_{k-1}|t_{k-1})$, the parameter prediction equation is used to predict the value of $\hat{\Theta}(t_k|t_{k-1})$. This constitutes a prediction of the state vector for $t=t_k$ based on the data up to $t=t_{k-1}$.
(2) The *innovation*, $\epsilon(t_k)$, is then computed as the difference between the new measurement taken at $t=t_k$ and the predicted measurement obtained by substituting $\hat{\Theta}(t_k|t_{k-1})$ into the measurement equation.
(3) Next, the Kalman gain or weight is computed based on this innovation.

TABLE II. Parameters of test cases. In all cases, the initial values of the parameters θ_1, θ_2, f_0, and δ are: 43°, $=-8°$, 50.1 Hz, and 2.5, respectively.

Case No.	No. of phones	SNR (dB)	Speed (m/s)	var(θ_1)(deg^2)	var(θ_2)(deg^2)
1	1	0	5	8.0×10^{-3}	5.0×10^{-3}
2	4	0	5	1.7×10^{-3}	3.8×10^{-3}
3	4	0	0	6.8×10^{-3}	6.9×10^{-3}
4	4	−10	5	5.8×10^{-3}	1.9×10^{-2}
5	4	−10	0	2.0×10^{-2}	3.0×10^{-2}
6	4	−10	0	4.5×10^{-2}	7.4×10^{-2}

(4) The Kalman gain is then used in the correction stage, producing $\hat{\Theta}(t_k|t_k)$, the corrected estimate.
(5) This corrected estimate is then substituted into the right-hand side of the prediction equation, thereby initiating the next iteration.

This completes the discussion of the model-based array processor. In the next section we will present some examples based on synthesized data.

III. PERFORMANCE ANALYSIS

In order to simplify the material in this section, we will present some examples based on a reasonably simple but nontrivial scenario. First we will assume that there are two plane-wave sources (see Fig. 1). Second, we will assume that the two sources are both operating at the same frequency. Finally, we will reduce the number of amplitude parameters, $\{a_m\}$ $m=1,2,...,M$, from two to one by rewriting Eq. (42) as

$$p(x_l,t_k)=a_1[e^{i\omega_1 t_k+\beta_l(t_k)\sin\theta_1}+\delta e^{i\omega_2 t_k+\beta_l(t_k)\sin\theta_2}]. \tag{52}$$

Here, $\delta=a_2/a_1$ and $\omega_1=\omega_2=\omega_0$. The parameter a_1 appearing outside the square brackets can be considered to be a data scaling parameter. Consequently, we have four parameters to deal with so that our measurement equation becomes

$$p(x_l,t_k)=e^{i\omega_0 t_k+\beta_l(t_k)\sin\theta_1}+\delta e^{i\omega_0 t_k+\beta_l(t_k)\sin\theta_2}, \tag{53}$$

and Eq. (46) simplifies to

$$\Theta=\begin{bmatrix}\theta_1\\ \theta_2\\ ---\\ \omega_0\\ ---\\ \delta\end{bmatrix}. \tag{54}$$

We now have all the necessary equations to implement the predictor/corrector form of the Kalman filter algorithm outlined in Table I. The calculations are carried out in MATLAB using the SSPACK_PC toolbox.[15]

In all of the following examples, we assume that the two sources are radiating narrow-band energy at a frequency of 50 Hz since this was one of the principal frequencies used in the Hudson Canyon experiment.[16] As outlined in Table II, the true values of the two bearings, θ_1 and θ_2, are 45 degrees and −10 degrees, respectively. The true amplitude ratio δ is 2. The corresponding initial values for θ_1, θ_2, $f_0=\omega_0/2\pi$, and δ are: 43°, −8°, 50.1 Hz, and 2.5, respectively.

Case 1. For this first example, we assume that there is a single hydrophone and everything is known *a priori* except the two bearings θ_1 and θ_2. This example is important since it demonstrates that indeed, the algorithm truly generates a synthetic aperture. The results are shown in Fig. 2, where the two bearing estimates are shown as a function of time. The signal-to-noise ratio (SNR) of the simulated signal is 0 dB for the unit amplitude signal, i.e., the signal associated with θ_1. Thus since $\delta=2$, the SNR of the signal associated with θ_2 is +6 dB. The duration of the signal in this example, as in all of the following examples, is 27 s. At 5 m/s, the hydrophone traces out an aperture of 135 m or 4.5 wavelengths (λ) at the specified frequency of $f_0=50$ Hz. Both bearing estimates converge to their true (mean) values and then ''walk'' about these means due to noise. In this and all subsequent figures, the dotted lines are the $2\times\sigma$ bounds as predicted by the algorithm.

Case 2. The number of hydrophones is increased to 4, the array's speed of forward motion is 5 m/s, and the SNR is 0 dB. This constitutes a 1.5 λ physical aperture and a synthetic aperture of 4.5 λ, a factor of 3 increase in aperture for

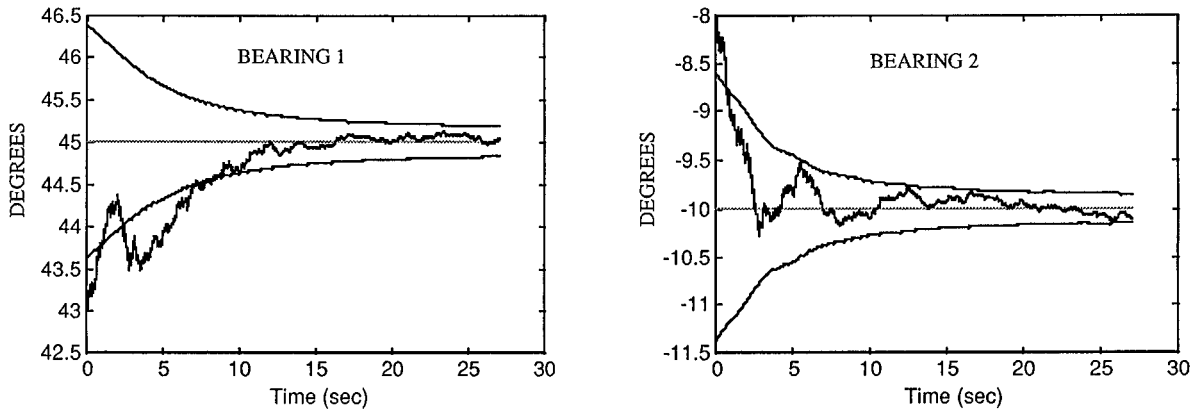


FIG. 2. Bearing estimates for case 1. Here there is a single hydrophone travelling at 5 m/s. All other parameters are assumed known *a priori*. The SNR for the signal associated with θ_1 is 0 dB. In this and all following figures, the dashed lines indicate the predicted $2\times\sigma$ bounds.

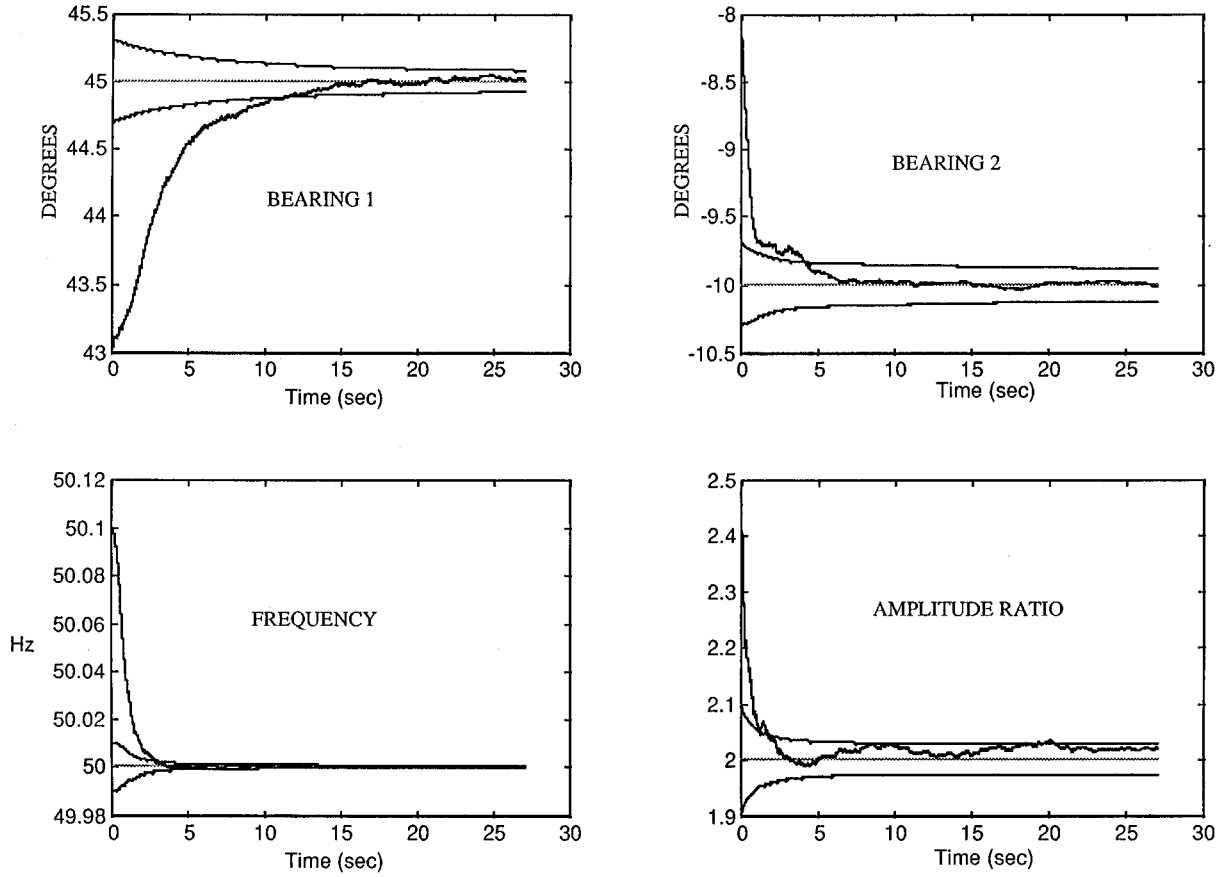


FIG. 3. Bearing, frequency, and amplitude estimates for case 2. Here there is a four hydrophone array traveling at 5 m/s. The SNR for the signal associated with θ_1 is 0 dB.

a total of 6 λ. The parameters being estimated are θ_1, θ_2, $f_0=\omega_0/2\pi$, and δ. The results are shown in Fig. 3 where we again see convergence to the true (mean) values. Here the limiting value of the variance is determined is the value of R_{ww}, the covariance assumed for the state process noise. This covariance plays the role of a tuning parameter in the Kalman filter algorithm and the smaller it is made, the more slowly the estimate converges.

Case 3. This is the same as the preceeding example except that the speed v has been set to zero. As can be seen in Fig. 4, the bearings do not converge as rapidly, since the synthetic effect is lost.

Case 4. This example is the same as case 2 except that the SNR has been decreased to -10 dB. From Fig. 5, it is seen that he estimates are noisier, as expected.

Case 5. In this example, we repeat case 4, but with the speed set to zero. In Fig. 6 we see, similarly with case 3, convergence is much slower due to the loss of the synthetic effect.

Case 6. In this final example, we show the effect of increasing R_{ww} in case 5. As mentioned above increasing this improves convergence, but at the cost of noiser estimates (Fig. 7).

IV. PERFORMANCE BOUNDS

An analysis of the performance bounds on a moving array has been treated in the work of Stergiopoulos[17] where it is shown that, if the source frequency is not known *a priori*, the Cramer–Rao lower bound (CRLB) on the variance of the bearing estimate is actually slightly worse (greater) than the case for the stationary array, but the *resolution* performance is improved. A more comprehensive treatment of the performance bounds on moving arrays has been given by Edelson[13] in which it is concluded that the CRLB on the bearing (source frequency) variance is lower when the source frequency (bearing) is known *a priori*. The variances of the estimates for our method, as predicted by the Kalman algorithm, are given in Table II where it is seen that, for the four hydrophone moving array, for example, where $v=5$ m/s, the variance on the estimate of θ_1, i.e., the bearing associated with the signal with SNR$=0$ dB (case 2), is 1.7×10^{-3} deg^2.

This result is in qualitative agreement with Edelson's analysis. Where, for the same case, his equation predicts a variance of 1.2×10^{-3} deg^2 since he assumes that there is only one source, where we have two. Also, in any case, the EKF cannot be guaranteed to yield an optimal minimum variance estimate. Another issue is that Edelson assumes that either the source frequency is known *a priori* or it is not, whereas we treat the source frequency *and* the two bearings as free parameters in a multivariate estimation problem. Nevertheless, our results are in qualitative agreement with Edelson's and clearly verify the fact that incorporating the motion of an acoustic array into the signal processing algorithm pro-

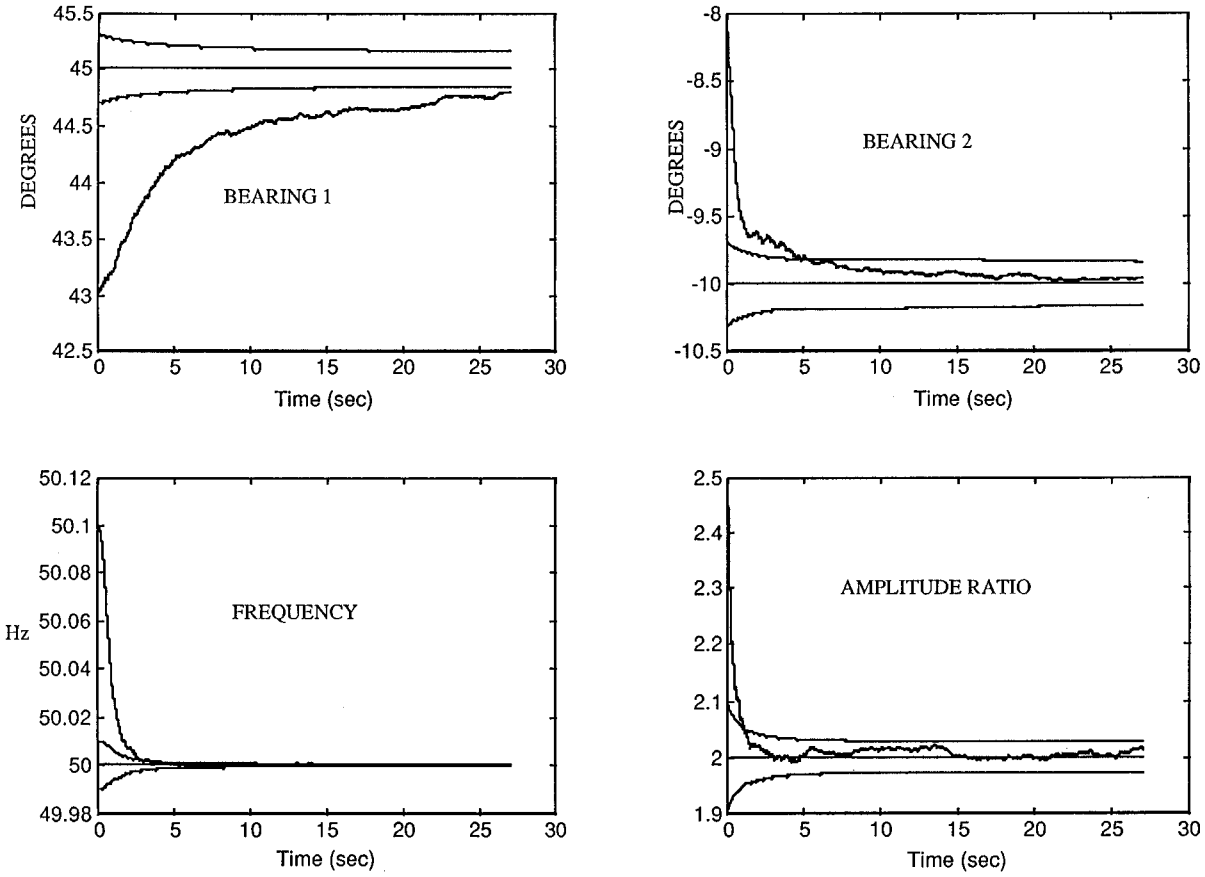


FIG. 4. Bearing, frequency, and amplitude estimates for case 3. Here there is a four hydrophone array travelling at 0 m/s. The SNR for the signal associated with θ_1 is 0 dB.

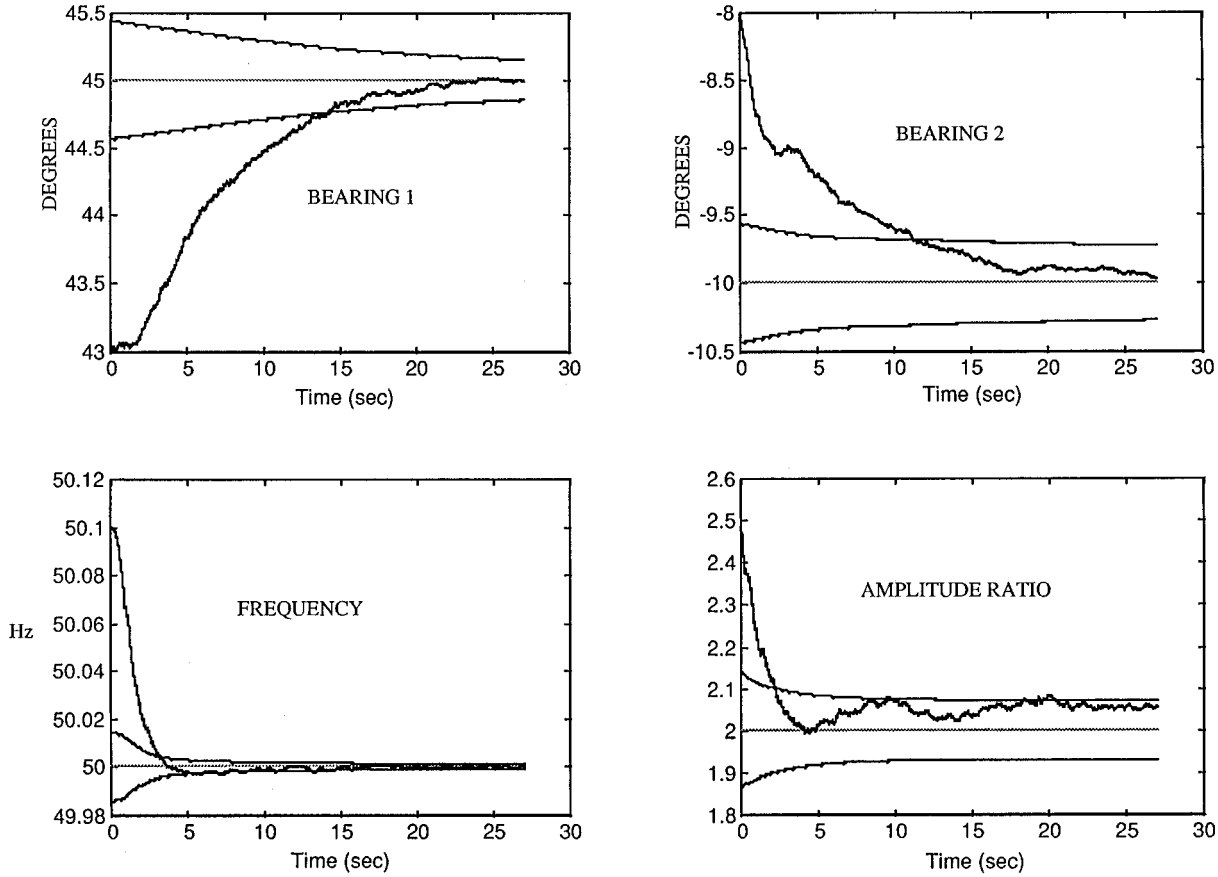


FIG. 5. Bearing, frequency, and amplitude estimates for case 4. Here there is a four hydrophone array with $v = 5$ m/s. The SNR for the signal associated with θ_1 is -10 dB.

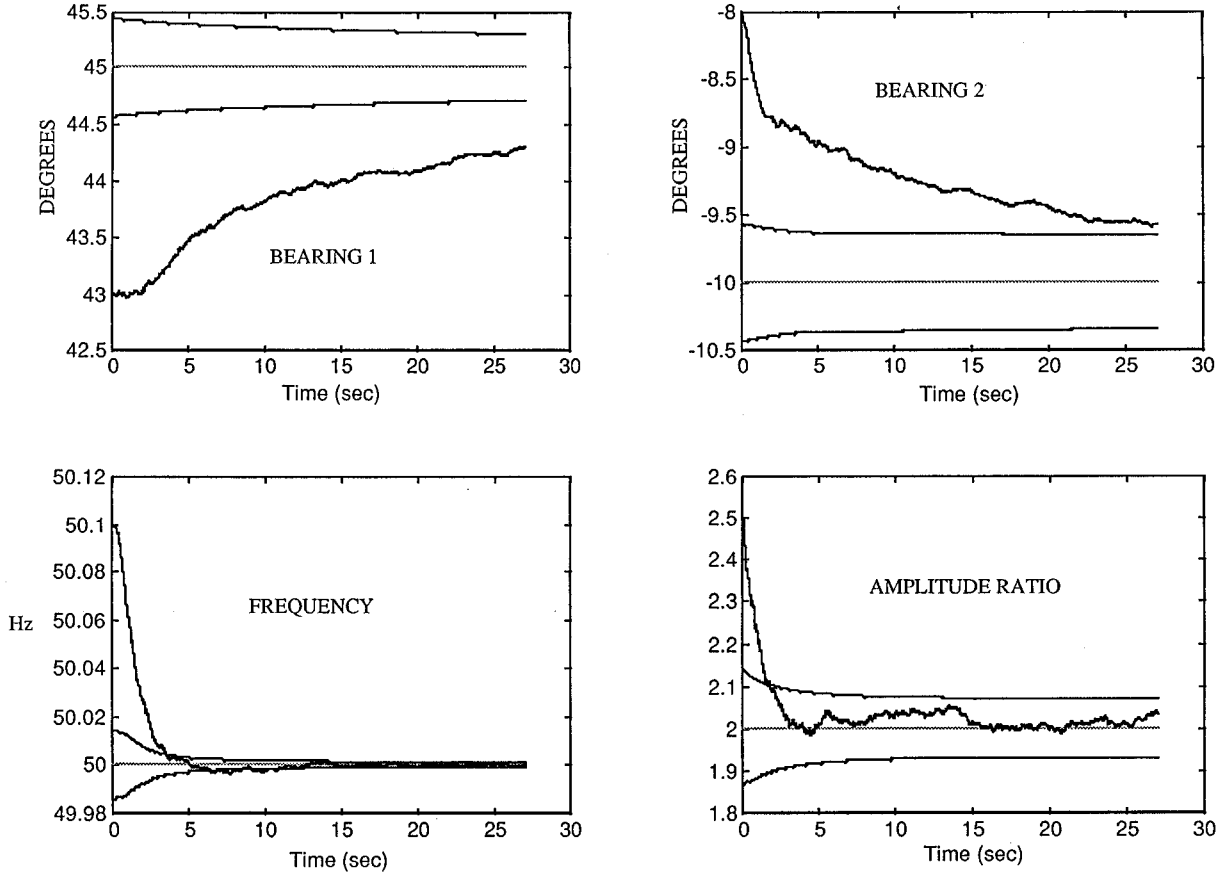


FIG. 6. Bearing, frequency, and amplitude estimates for case 5. Here there is a four hydrophone array travelling at 0 m/s. The SNR for the signal associated with θ_1 is -10 dB.

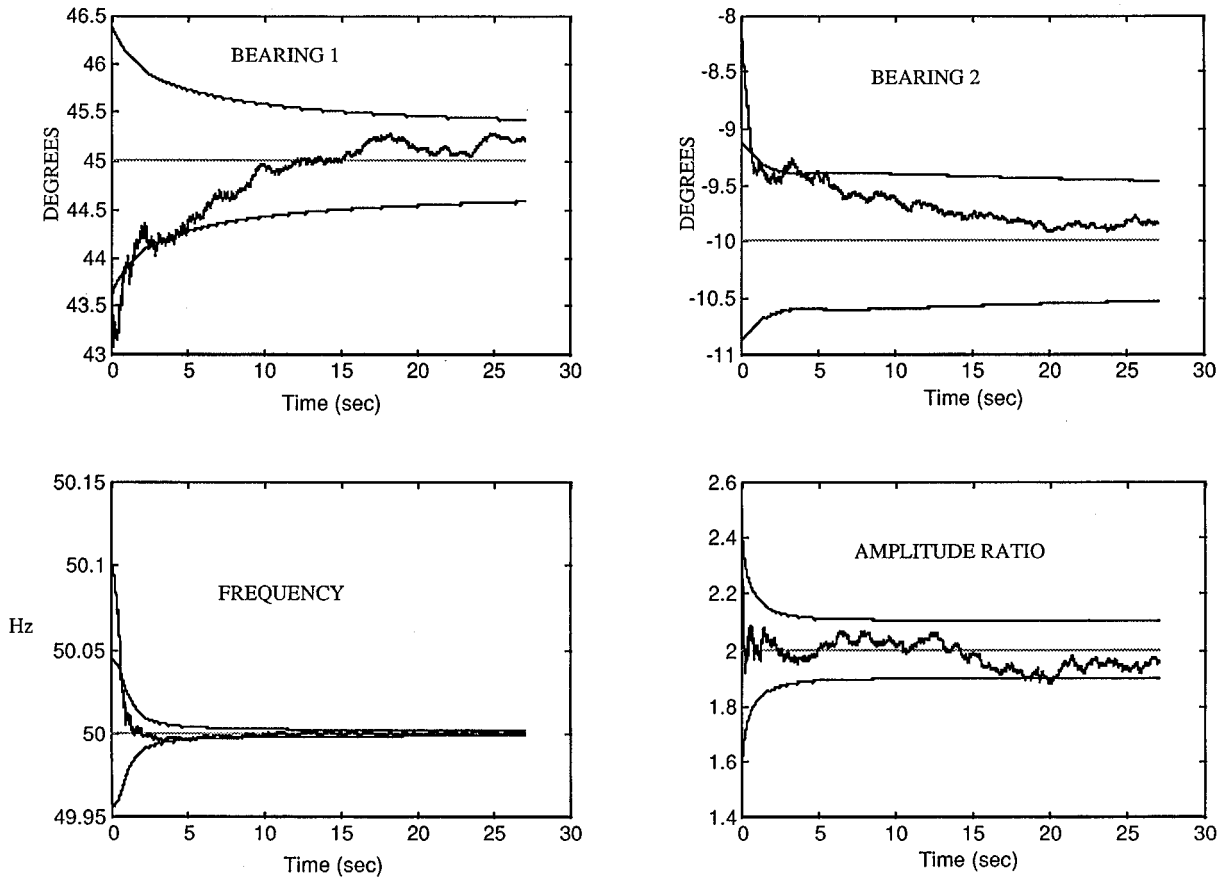


FIG. 7. Bearing, frequency, and amplitude estimates for case 5, with the value of the system noise covariance increased to improve the convergence.

duces a significant improvement in performance. This is also borne out in Table II where it is seen that when the array motion is set to zero (cases 3 and 5), the predicted variance increases significantly over the corresponding cases (cases 2 and 4), where the array is moving.

These results emphasize the point that the CRLB is only attained by an *efficient* estimator, which means that the results are clearly dependent on the model and algorithm chosen for a model-based estimation scheme. Standard acoustic array bearing estimators (beamformers) do not take advantage of the forward motion of the array and therefore cannot comprise efficient estimators of the source bearing(s). A popular form of bearing estimator is the so-called k-ω beamformer, wherein the beamformer takes the form of a discrete spatial Fourier transform. This introduces a limit on the bearing accuracy, since the finite bin size of the spatial transform limits the bearing resolution, unless the size of the transform can somehow be increased in an adaptive manner.

V. DISCUSSION AND CONCLUSIONS

The model-based approach has been applied to space–time acoustic array processing. By explicitly including the motion of the array in the signal model, improved bearing estimation and resolution performance is obtained. The technique is shown to be equivalent to a passive synthetic aperture processor which allows the motion of the array to effectively increase its useful aperture, thereby providing all of the associated improvement in performance. The advantage of this approach is that there is essentially no limit to the degree of sophistication allowed for the particular models chosen for the processor. In this work we have chosen the signal model to be a sum of plane waves. However, the method can be easily generalized to more general models such as signals with spherical or cylindrical wavefronts which, for example, would permit a wavefront curvature ranging scheme to be implemented that could exploit the large apertures available. The introduction of more sophisticated propagation models such as the normal mode model and the parabolic equation model would be an interesting next step.[2,18] The richness of these models permits ranging when the array is in or near the endfire mode.

As mentioned before, an interesting aspect of this processor is that it performs bearing estimation without the necessity of *introducing* an explicit beamformer structure. The advantage of this is that all of the limitations of standard beamformers, such as finite beam bin sizes and the limitation to a fixed number of beams, are avoided. A further advantage of this approach is that the innovation sequence provided by the Kalman estimator carries information regarding the performance of the model. For example, if we had assumed fewer or more signal sources than there actually were, the innovations would no longer be zero-mean and white, indicating a mismatch of the model to the physics.[7] Although we have not exploited this here, there remains the potential of using this information to monitor and update the model online. Thus deviations from plane-wave signals and distortions due to array to motion could, in principle, be compensated for in a self-consistent manner.

This work was limited to the multiple narrow-band frequency case. A logical next step would be to generalize the processor to include the broadband case. This would require a broadband signal model which could, for example, be modeled as an ARMA process where the parameters of the ARMA model could be augmented directly into the state vector.

Finally, we would like to point out that a great deal of work on target tracking using Kalman filter processing of acoustic array data as been carried out in the past. This work differs from the work described here in that these methods utilize the bearing estimates from a standard beamformer as the measurements for the purpose of estimating the coordinates of a target. For more information on these techniques see Ref. 19 and references therein.

ACKNOWLEDGMENTS

The authors would like to acknowledge helpful conversations with Dr. R. Pridham of the Raytheon Company who, along with Drs. J. Bartram and W. Knight, considered a similar approach to array processing in an unpublished report. This work was funded by the Advanced Projects Research Agency under ARPA Order No. A695.

[1] E. J. Sullivan and J. V. Candy, ''Passive synthetic aperture processing as a Kalman filter problem,'' J. Acoust. Soc. Am. **95**, 2953 (1994).
[2] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics*, AIP Series in Modern Acoustics and Signal Processing (AIP Press, New York, 1994).
[3] M. J. Hinich, ''Maximum likelihood signal processing for a vertical array,'' J. Acoust. Soc. Am. **54**, 499–503 (1973).
[4] H. P. Bucker, ''Use of calculated sound fields and matched-field detection to locate sound sources in shallow water,'' J. Acoust. Soc. Am. **59**, 368–372 (1976).
[5] R. D. Doolittle, A. Tolstoy, and E. J. Sullivan, ''Editorial in special issue on detection and estimation in matched-field processing,'' IEEE J. Ocean Eng. **18**, 153–155 (1993).
[6] J. V. Candy and E. J. Sullivan, ''Model-based identification: an adaptive approach to ocean-acoustic processing,'' IEEE J. Ocean Eng. **21**, 273–289 (1996).
[7] J. V. Candy, *Signal Processing: The Model-Based Approach* (McGraw-Hill, New York, 1986).
[8] A. Tolstoy, *Matched Field Processing for Ocean Acoustics* (World Scientific, River Edge, NJ, 1993).
[9] R. E. Williams, ''Creating an acoustic synthetic aperture in the ocean,'' J. Acoust. Soc. Am. **60**, 60–73 (1976).
[10] N. Yen and W. Carey, ''Application of synthetic aperture processing to towed array data,'' J. Acoust. Soc. Am. **60**, 764–765 (1976).
[11] S. Stergiopoulos and E. J. Sullivan, ''Extended towed array processing by an overlapcorrelator,'' J. Acoust. Soc. Am. **86**, 158–171 (1989).
[12] E. J. Sullivan, W. M. Carey, and S. Stergiopoulos, ''Editorial in special issue on acoustic synthetic aperture processing,'' IEEE J. Ocean Eng. **17**, 1–7 (1992).
[13] G. S. Edelson, ''On the Estimation of Source Location Using a Passive Towed Array,'' Ph.D. dissertation, University of Rhode Island, 1993.
[14] L. J. Ljung, *System Identification: Theory for the User* (Prentice-Hall, Englewood Cliffs, NJ, 1987).
[15] J. V. Candy and P. M. Candy, ''SSPACK_PC, A model-based signal processing package on personal computers,'' DSP Applic. **2**, 33–42 (1993).
[16] L. Maiocco, W. Carey, E. Parssinen, and J. Doutt, ''Measurement of shallow water sound transmission on the New Jersey continental shelf,'' J. Acoust. Soc. Am. Suppl. 1 **86**, S8 (1989).

[17] S. Stergiopoulous, ''Optimum bearing resolution for a moving towed array and extention of its physical aperture,'' J. Acoust. Soc. Am. **87**, 2128–2140 (1990).
[18] A. R. Robinson and D. Lee, *Oceanography and Acoustics: Prediction and Propagation Models*, AIP Series in Modern Acoustics and Signal Processing (American Institute of Physics, New York, 1994).
[19] J. C. Hassab, *Underwater Signal and Data Processing* (CRC, Boca Raton, FL, 1989).

The generation of psychoacoustic combination tones in relation to two-tone suppression effects in a computational model

Christian Giguère[a)]
Audiology/Speech-Language Pathology Program, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada

Guido F. Smoorenburg
Laboratory of Experimental Audiology, Department of Otorhinolaryngology, University Hospital Utrecht, 3508 GA Utrecht, The Netherlands

Hans Kunov
Institute of Biomedical Engineering, University of Toronto, Toronto, Ontario M5S 3G9, Canada

(Received 12 September 1996; revised 2 July 1997; accepted 7 July 1997)

The amplitude of the psychoacoustic distortion product $2f_1-f_2$ elicited by primaries f_1 and f_2 depends on the method of measurement. The cancellation-tone procedure gives consistently higher estimates of the distortion product than nonsimultaneous procedures. It was suggested that the difference could be attributed to suppression effects by the lower primary f_1 [Smoorenburg, J. Acoust. Soc. Am. **52**, 615–632 (1972)]. Simulations carried out with a computational model of the auditory periphery including a single, but distributed, compressive nonlinearity confirm the hypothesis; the amount of suppression produced at $2f_1-f_2$ by f_1 equals the difference between simultaneous and nonsimultaneous methods of measuring the amplitude of the distortion product. This is in agreement with psychoacoustical data [Shannon and Houtgast, J. Acoust. Soc. Am. **68**, 825–829 (1980)]. The model also reveals the importance of suppression effects by f_1 to bring the phase of the traveling wave associated with the cancellation tone $2f_1-f_2$ in exact opposition with that of the distortion product $2f_1-f_2$ generated by the primaries. There is also evidence of suppression effects by the higher primary f_2 when $L_2 \gg L_1$. In the model, cancellation occurs over a substantial portion of the basilar membrane near the characteristic place for $2f_1-f_2$. [S0001-4966(97)03311-0]

PACS numbers: 43.64.Bt, 43.64.Jb, 43.66.Ba, 43.66.Ki [RDF]

INTRODUCTION

Auditory distortion products of the type $f_1-n(f_2-f_1)$ are generated by the human ear when stimulated by two closely spaced pure tones f_1 and f_2, where $f_1<f_2$. The strongest member from this set, $2f_1-f_2$ ($n=1$), has traditionally been studied as a combination tone (CT) measured subjectively in cancellation-tone, loudness-matching, or detection-threshold psychoacoustic experiments (Zwicker, 1955, 1968; Goldstein, 1967; Helle, 1969; Smoorenburg, 1972a, b). More recently, it has been studied mostly as a distortion-product otoacoustic emission (DPOAE) measured objectively in the ear canal using sensitive electroacoustic equipment (Kemp, 1979; Probst *et al.*, 1991). Both CT and DPOAE phenomena are widely believed to have a common cochlear origin and appear to be closely related to the functioning of the outer hair cells (Probst *et al.*, 1991). In addition to the clinical diagnostic potential of these phenomena, their study can bring important new information about the active and nonlinear mechanisms of the cochlea. Currently, however, there is a less than complete understanding of the exact relation between CTs and DPOAEs, and between the different methods of measuring CTs. This study focuses on the latter aspect.

Two main psychoacoustic methods have been used to measure the growth of the perceived CTs as a function of the levels L_1 and L_2 of the two stimulus tones. In the cancellation-tone method (Goldstein, 1967; Smoorenburg, 1972b, 1974), a probe tone is presented externally and simultaneously with the primary stimulus tones f_1 and f_2 at the frequency of the CT to be measured. The task of the subject is to adjust the probe amplitude and phase until the pitch sensation of the CT becomes inaudible. The common interpretation of this method is that, at the cancellation point, the probe tone has the same amplitude as the internally generated CT but is 180° out-of-phase. Measured in this way, the CT amplitude grows with stimulus level for $L_1=L_2$ with a slope close to 1.0 dB/dB at low and moderate stimulus levels, and tends to saturate gradually at high levels. The CT amplitude also grows proportionally with L_2 for L_1 = constant until L_2 approaches L_1, where it peaks, and then decreases with further increase in L_2 with a slope that is subject dependent.

A second psychoacoustic method consists of matching the loudness of the CT with that of a probe tone of the same frequency presented externally but nonsimultaneously with the primary stimulus tones f_1 and f_2, e.g., by use of the pulsation-threshold technique (Smoorenburg, 1972b, 1974;

[a)]Also at Laboratory of Experimental Audiology, Department of Otorhinolaryngology, University Hospital Utrecht, 3508 GA Utrecht, The Netherlands, Electronic mail: c.giguere@med.ruu.nl and g.smoorenburg@kmb.azu.nl

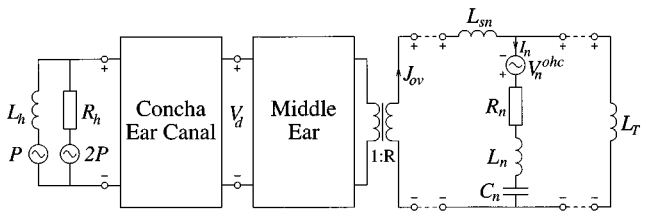


FIG. 1. Electroacoustic representation of the model.

Houtgast, 1972; Shannon and Houtgast, 1980). This method is referred to here as the comparison-tone method. Measured in this way, the growth of the CT amplitude with the levels L_1 and L_2 of the two stimulus tones shows the same characteristic patterns as those observed via the cancellation-tone method. However, the comparison method gives consistently lower estimates of the CT level than the cancellation method. The difference between the two methods increases with increasing stimulus level for $L_1=L_2$ and decreases with increasing frequency ratio f_2/f_1. When the CT amplitude is measured as function of L_2 for $L_1=$constant, the difference between methods is essentially constant, so that level L_1 seems to be the controlling factor. It has been suggested that the difference between cancellation and comparison methods can be attributed to suppression effects by f_1 (Smoorenburg, 1972b). However, mathematical analyses of the behavior of memoryless nonlinearities and several psychoacoustic experiments conducted to establish the exact link between CT generation and suppression effects were not fully conclusive (Smoorenburg, 1972b, 1974). Later, Shannon and Houtgast (1980) have been able to confirm in three out of four subjects that the difference between cancellation-tone and pulsation-threshold measurements coincides with the amount of suppression produced by the lower primary f_1 on a tone at the frequency of the CT. On the other hand, Humes (1983) tentatively concluded on the basis of the slopes of the suppression and combination-tone growth functions that these two phenomena are not directly related and cannot be described by one and the same nonlinearity.

The main objective of this paper is to study, within a computational modeling framework, the difference between cancellation-tone and comparison-tone methods of measuring CTs. In particular, the role of suppression effects by f_1 is explored. The computational model of the auditory periphery described in Giguère and Woodland (1994) is used. The cochlear stage includes distributed nonlinear elements derived from the fast motile responses of the outer hair cells (OHCs), and this serves as the basis for generating distortion products. The computer simulations suggest that suppression effects on the cancellation tone by stimulus f_1 play a key role in determining the difference between cancellation and comparison methods of measuring the $2f_1-f_2$ CT. Only when the amplitude reduction and phase shift of the cancellation-tone traveling wave due to f_1 are accounted for is the cancellation tone in exact amplitude and opposite phase with the internally generated CT.

I. MODEL

The peripheral auditory model (Giguère and Woodland, 1994) can be represented as a lumped-parameter electroacoustic network as shown in Fig. 1. The elements L_h and R_h and sources $P(t)$ and $2P(t)$ form an equivalent circuit for an acoustic signal laterally incident upon the human head including sound diffraction effects. The incident pressure wave $P(t)$ is then processed through a series of cross-coupled outer ear, middle ear, and cochlear stages. The outer ear stage accounts for the sound propagation through the concha and ear canal. The middle ear stage is based on the model of Lutman and Martin (1979) modified to provide a bidirectional coupling with the cochlear stage. The time-variant capacitive element used to model the effect of stapedial muscle contractions (Giguère and Woodland, 1994) is not used in this study to keep the initial stages of the model linear.

The cochlear stage is based on a nonlinear one-dimensional transmission line approximation of basilar membrane (BM) motion, discretized into $N=320$ channels from base ($n=1$) to apex ($n=N$). The circuit elements R_n, L_n, and C_n are analogous to the effective acoustic resistance, mass and compliance of the cochlear partition in channel n. The natural frequency $f_n=1/2\pi(L_nC_n)^{-0.5}$ of this second-order shunt resonant system decreases from base to apex according to the human cochlear place-frequency map of Greenwood (1990). The inductor L_{sn} in each channel represents the acoustic mass of the scalae vestibuli and tympani fluids. The apical end of the transmission line is terminated by an inductor L_T representing the acoustic mass of cochlear fluid from the last BM channel to the helicotrema. The voltage $V_d(t)$ and current $J_{0v}(t)$ signals correspond to the eardrum sound pressure and stapes volume velocity, respectively. The current $I_n(t)$ represents the BM volume velocity.

In each BM channel, a voltage source $V_n^{\mathrm{ohc}}(t)$ produces a nonlinear and saturating pressure, assumed to originate from the OHCs, as follows:

$$V_n^{\mathrm{ohc}}(t)=G(I_n(t))*R_n^*I_n(t), \tag{1a}$$

where:

$$G(I_n(t))=G^*(1+|I_n(t)|/I_0)^{-s}, \tag{1b}$$

and I_0 is an input scaling constant (3.6×10^{-3} cm/s times the BM channel cross-sectional area), G is the maximum gain, and s is a power exponent. Equation (1) is meant to embody the bidirectional electrophysiological transduction processes of the OHCs (Patuzzi *et al.*, 1989) within a simple relation. Functionally, the voltage source $V_n^{\mathrm{ohc}}(t)$ compensates for the pressure loss across the BM resistance R_n and thus reduces the damping of the BM. $V_n^{\mathrm{ohc}}(t)$ saturates gradually with increasing BM vibration through the nonlinear gain $G(I_n(t))$. The shape of this nonlinear gain is based on previous psychoacoustic and modeling studies of CT generation (Smoorenburg, 1972b, 1974; Duifhuis, 1976; Strube, 1986) which indicated a signal-compressing type of nonlinearity such as a power law with exponent between 0 and 1, typically around 0.6. However, a pure power-law function with exponent between 0 and 1 could lead to instability problems in a time-domain model due to a singularity and expansive behavior at zero BM vibration [$I_n(t)=0$]. To avoid this problem, a constant term is included within the power function in Eq. (1b), which has the effect of linearizing the model

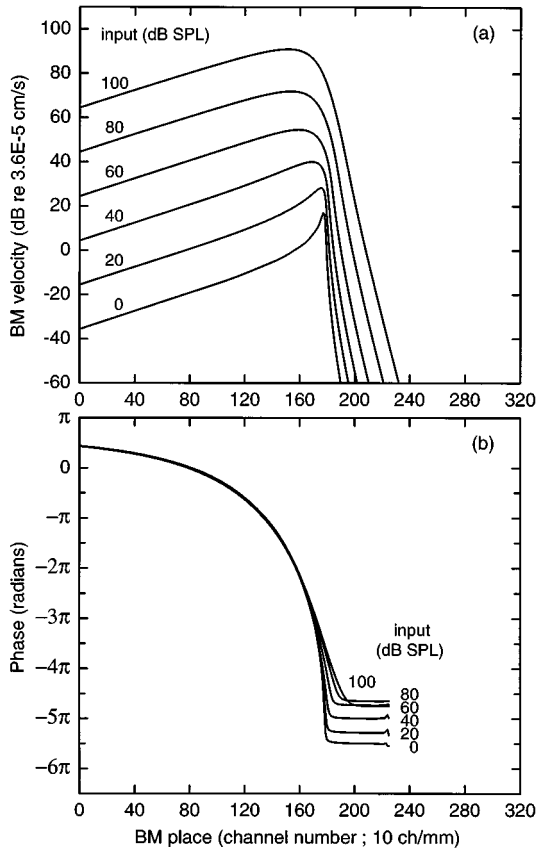


FIG. 2. BM velocity response (a) amplitude and (b) phase as a function of place from base ($n=1$) to apex ($n=320$) for a pure-tone stimulus of 1120 Hz at several input levels.

$[G(I_n(t))=G]$ at low BM vibration amplitudes $[I_n(t) \ll I_0]$. At higher amplitudes $[I_n(t) \gg I_0]$, the nonlinearity reduces to the pure power-law function.

In the model, $V_n^{\mathrm{ohc}}(t)$ depends only on the local BM volume velocity $I_n(t)$ at place n without direct lateral coupling nor an additional stage of filtering or phase shift. The maximum gain is also uniform across the cochlea from base to apex and maintained at $G=0.99$, giving a residual BM damping of at least 1% of that of the linear passive version of the model ($G=0$). Although lateral coupling and nonuniform gains along the cochlea could be important factors to consider in explaining the fine frequency structure and other intricate details of auditory distortion products (Sun *et al.*, 1994a, b; Fukasawa and Tanaka, 1994), the present study is mainly concerned with the gross level behaviour of the combination tones. The implications of a different gain and of filtering of the feedback signal are discussed in Sec. V.

Numerically, the model is implemented by translating the electroacoustic representation into a topologically equivalent time-domain computational structure using the technique of wave digital filtering (Fettweis, 1986). The sampling rate of operation of the computational model is 71 680 Hz. Preliminary results were found to be relatively insensitive to the exponent s in Eq. (1b) for values around 0.4–0.6. A value of $s=0.5$ was thus chosen throughout the paper in view of the substantial computational advantage of implementing the power function as a square root. All other model parameters are chosen as in Giguère and Woodland (1994).

Figure 2(a) shows the excitation pattern on the BM of the model for a steady-state pure-tone stimulus $P(t)$ of frequency 1120 Hz presented at several input levels. The excitation pattern was computed as the root-mean-square average of the BM velocity (i.e., $I_n(t)$ divided by the BM channel cross-sectional area) in every channel n of the cochlea and scaled in dB using an arbitrary velocity reference of 3.6×10^{-5} cm/s. In Fig. 2(a), the excitation along the BM increases gradually from the basal end of the cochlea to the place of resonance and decreases rapidly afterward. At low input levels, the BM is sharply tuned. At high input levels, the BM is much more broadly tuned and the place of resonance moves basally by an amount equivalent to a shift of about 1/2 octave on the cochlear place-frequency map. The same amount of shift is reported by Johnstone *et al.* (1986) for the guinea pig. Near the characteristic place, the BM vibration grows at a lower rate than the stimulus. Over the full input range, there is about 43 dB of BM level compression at the low-level characteristic place ($n=177$) and 31 dB along the locus of maximum vibration. Corresponding values of 55 and 40 dB can be inferred from the data reported in Johnstone *et al.* (1986). The bandwidths of the response peaks in the model are in close agreement with the physiological data from Johnstone *et al.* (1986) for equivalent input levels, although other and more recent physiological data (Robles *et al.*, 1986; Ruggero, 1992) tend to indicate less variation with input level.

Figure 2(b) shows the phase behavior of the stimulus tone along the BM. The phase value at the basal end of the cochlea is the phase shift arising from the transmission characteristics of the outer ear and middle ear at 1120 Hz. The phase of the traveling wave then decreases gradually from the base to the place of resonance for 1120 Hz, where it drops suddenly, and reaches a plateau thereafter.

II. COMBINATION TONES

A. Generation of CTs

Figure 3 presents an example of the BM velocity excitation pattern for a two-tone stimulus $f_1=1400$ Hz and $f_2=1680$ Hz, which corresponds to $f_2/f_1=1.20$, presented at several input levels for the case $L_1=L_2$. The stimulus tones f_1 and f_2 are generated as cosine waveforms with $\phi_1=\phi_2=0°$. In Fig. 3, the first two peaks from the base of the cochlea correspond to the resonance places for f_2 and f_1 respectively. In the overlapping region of their excitation patterns, combination tones of the type $f_1-n(f_2-f_1)$, $n=1,...,4$, are generated. These four CTs have lower frequencies [1120 Hz ($n=1$), 840 Hz ($n=2$), 560 Hz ($n=3$), and 280 Hz ($n=4$)] than the two stimulus tones f_1 and f_2, and thus travel further toward the apex and are resolved at their own place of resonance. The existence of first- and higher-order CTs has been found both psychoacoustically (Smoorenburg, 1972a) and electrophysiologically (Smoorenburg *et al.*, 1976). In Fig. 3, the most prominent peak ($n=1$) corresponds to the commonly studied $2f_1-f_2$ auditory distortion product. At low input levels (20–40 dB SPL), the two stimulus tones and all four CTs are resolved as clear

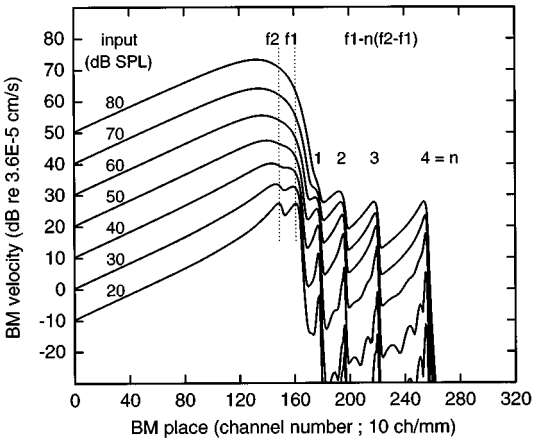


FIG. 3. BM excitation pattern for a two-tone stimulus $f_1=1400$ Hz and $f_2=1680$ Hz at several input levels $L_1=L_2$ showing the generation of combination tones of the type $f_1-n(f_2-f_1)$. Note: dotted lines indicate the position of the low-level characteristic place corresponding to f_2 and f_1.

peaks. As the level increases (60–80 dB SPL), the excitation patterns due to f_1 and f_2 greatly overlap near their place of resonance where they give rise to a single relatively broad peak. The apical tail of the excitation pattern due to f_1 also tends to override the first CT. At 80 dB SPL, the $2f_1-f_2$ CT is almost completely overridden by f_1.

Combination tones of the type $f_2-m(f_1-f_2)$, $m\geqslant 1$, are also generated by the model but are not resolved in Fig. 3. They have frequencies higher than the two stimulus components and thus are completely overridden by the excitation patterns due to f_1 and f_2. This is in agreement with Smoorenburg *et al.* (1976) who could not find trace of CTs above the stimulus frequencies in electrophysiological recordings. These products can be detected as DPOAEs, however, from spectral analysis of the eardrum sound pressure both in the model and in real ears (Giguère *et al.*, 1995a, b).

B. Cancellation-tone method

In the cancellation-tone psychoacoustic experiment, a probe tone at the frequency of the CT to be measured is presented externally and simultaneously with the primary stimulus tones f_1 and f_2. The task of the subject is to cancel the pitch sensation of the CT by adjusting the amplitude and phase of the probe tone. Cancellation-tone experiments are reproduced in the model by adjusting the amplitude and phase of a probe tone until the peak in the BM excitation pattern at the corresponding CT frequency disappears completely. This is illustrated in Fig. 4(a) for a cancellation-tone simulation at $L_1=L_2=60$ dB SPL, and frequencies f_1 and f_2 as in Fig. 3. The thick line is the excitation pattern when the primary stimulus tones f_1 and f_2 are presented alone. The thin line is the excitation pattern when a cancellation probe tone f_3 at the frequency of the first combination tone ($2f_1-f_2=1120$ Hz) is added to the primary tones. As can be seen, the excitation pattern corresponding to the first combination tone disappears completely so that the tail of the excitation pattern due to f_1 extends much further toward the apex up to the second combination tone. Electrophysiological evidence for the removal of nerve discharges tuned to the

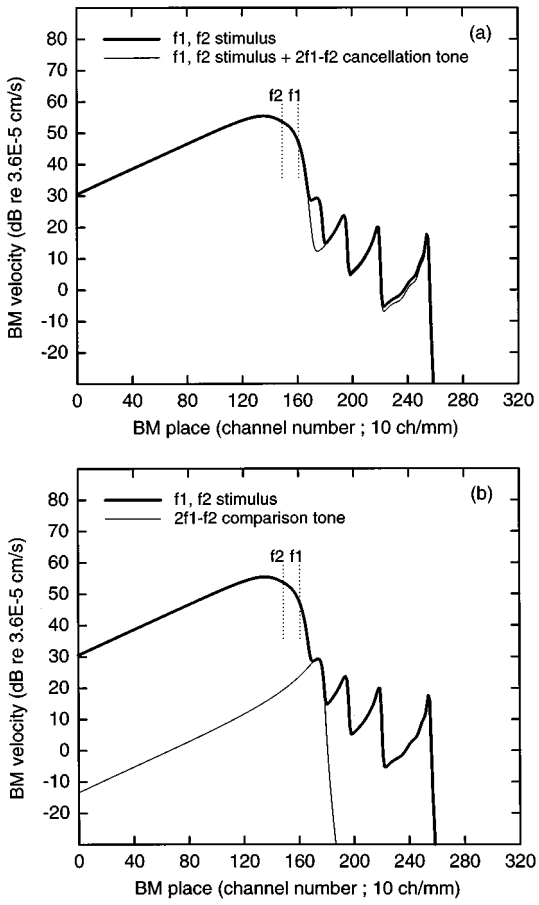


FIG. 4. BM excitation pattern for a two-tone stimulus $f_1=1400$ Hz and $f_2=1680$ Hz presented at $L_1=L_2=60$ dB SPL (thick line), and for the simulation of the $2f_1-f_2$ CT measurement (thin line) using the (a) cancellation-tone method ($L_3=31.4$ dB SPL, $\phi_3=-17°$) and (b) comparison-tone method ($L_3=21.8$ dB SPL, $\phi_3=192°$). Note: dotted lines indicate the position of the low-level characteristic place corresponding to f_2 and f_1.

CT by an external cancellation tone of the same frequency was reported by Goldstein (1970). In Fig. 4(a), the excitation patterns from the two primary tones f_1 and f_2, and from the remaining three CTs are essentially unaffected by the cancellation tone. In this example, the level and phase of the probe tone required for cancellation are $L_3=31.4$ dB SPL and $\phi_3=-17°$, respectively. As in psychoacoustic experiments, the model adjustments are quite sensitive to small variations in the level and phase of the cancellation tone, so that in general this procedure leads to a clear cancellation point. Typical simulation accuracy in this study is better than 0.2 dB and 2°.

C. Comparison-tone method

Broadly speaking, the comparison-tone psychoacoustic method consists of matching the perceived loudness of the CT with that of a probe tone of the same frequency presented externally but nonsimultaneously. A typical realization of this method is by use of the pulsation-threshold alternation technique, initially developed by Houtgast (1972) for lateral inhibition studies, and applied to CT measurements by Smoorenburg (1972b, 1974). Essentially, the two-tone stimulus f_1 and f_2 generating the $2f_1-f_2$ CT is alternated at a

rate of 4 Hz with a probe tone of frequency $f_3=2f_1-f_2$ using complementary on/off gates, so that an alternation of the same stimuli would produce a continuous signal. In these experiments, the phase of the comparison probe tone is set 180° out-of-phase with that of the cancellation tone for the given stimulus f_1 and f_2. The assumption is that the internally generated CT, and thus the comparison tone, are in phase opposition with the cancellation tone. The level of the probe is then increased until a pulsation of this tone becomes just audible. Just below this threshold a continuous tone is perceived, and this is interpreted as the matching of the levels of the probe tone and the internally generated CT produced by the stimulus f_1 and f_2. No independent estimate of the CT phase is obtained using this method. An alternative nonsimultaneous method based on gap masking yields essentially the same results (Smoorenburg, 1972b).

Comparison-tone experiments are easily reproduced in the model under the assumption that, at the matching point, the BM excitation pattern due to the probe tone is equal to that of the CT during the f_1 and f_2 stimulus presentation. The matching point is chosen as the BM place whose characteristic frequency equals that of the CT frequency. In other words, the simulation of the comparison-tone experiment consists of finding the curve from the set of Fig. 2(a) that best matches the peak of the CT in Fig. 3. This is illustrated in Fig. 4(b) for the same two-tone stimulus as in Fig. 4(a). The thick line is again the excitation pattern when the primary stimulus tones are presented alone. The thin line is the excitation pattern for a single tone $f_3=2f_1-f_2=1120$ Hz that best matches the peak of the first CT. As can be seen in Fig. 4(b), the comparison-tone simulation leads in general to a match of not only the absolute peak but also of the whole tip region of the CT excitation pattern. A comparison phase value can also be obtained with the model if the traveling phases of the comparison tone and the CT are to be matched in the tip region. In some conditions, the CT does not lead to a clear peak on the BM excitation pattern. An example of this case was previously noted in Fig. 3 at high stimulus levels, where the $2f_1-f_2$ CT is detectable only as a disturbance to the steeply decreasing excitation pattern from the f_1 primary. The comparison-tone simulation consisted then of adjusting the amplitude and phase of the probe tone to just match the residual excitation pattern from the CT at the BM place for $2f_1-f_2$. Typical simulation accuracy for the comparison method is identical to the cancellation method.

In the example of Fig. 4(b), the level and phase of the matching probe tone are $L_3=21.8$ dB SPL and $\phi_3=192°$, respectively. Thus in general, the model cancellation and comparison levels are different (31.4 and 21.8 dB SPL, respectively), and cancellation phase ($-17°$) is not necessarily 180° out-of-phase with comparison phase (192°).

D. Results

1. Case $L_1=L_2$

Figure 5(a) presents the growth of the $2f_1-f_2$ CT in the model with increasing stimulus level $L_1=L_2$ for the same two-tone stimulus $f_1=1400$ Hz and $f_2=1680$ Hz ($f_2/f_1=1.20$) as in Figs. 3 and 4. The results from both

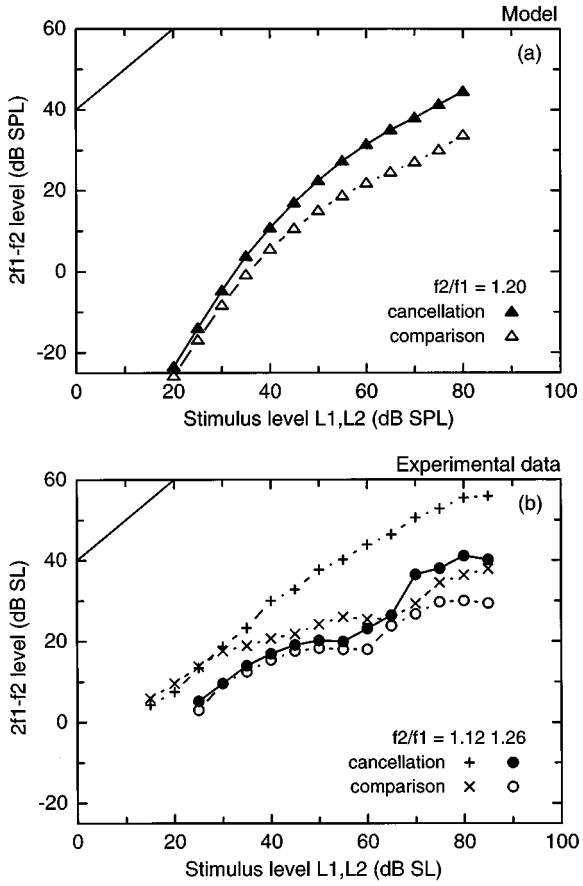


FIG. 5. Growth of the $2f_1-f_2$ combination tone with stimulus level for the case $L_1=L_2$: (a) model results for $f_1=1400$ Hz and $f_2/f_1=1.20$; (b) experimental results for one subject (GS) from Smoorenburg (1974) for $f_1=1000$ Hz, and $f_2/f_1=1.12$ and 1.26.

measurement methods are shown. The $2f_1-f_2$ CT increases monotonically with stimulus level with either method. Overall, the slope of the cancellation levels is close to 1.0 dB/dB and tends to decrease slightly at higher stimulus levels, in agreement with psychoacoustical data (Smoorenburg, 1972b, 1974). The model comparison levels are lower than the model cancellation levels and tend to saturate faster at high stimulus levels, as also found psychoacoustically (Smoorenburg, 1974; Shannon and Houtgast, 1980). Figure 5(b) presents psychoacoustical data from Smoorenburg (1974) for one subject for $f_1=1000$ Hz and two frequency ratios $f_2/f_1=1.12$ and 1.26. The comparison levels were obtained using the pulsation-threshold technique.

The model comparison and cancellation levels are in good agreement with the psychoacoustical data at mid to high stimulus levels and a frequency ratio f_2/f_1 of about 1.20 (Fig. 5; also Smoorenburg, 1972b). However, the slopes of the model comparison and cancellation levels [Fig. 5(a)] are slightly higher than 1.0 dB/dB at low stimulus levels ($\leqslant 40$ dB SPL), typically 1.5 dB/dB or more due to the linearization of the model for $I_n(t)\ll I_0$, and thus the psychoacoustical data [Fig. 5(b); also Smoorenburg, 1972b] tend to be underestimated by 5–15 dB in this region.

The difference between cancellation and comparison levels increases with stimulus level for both model and ex-

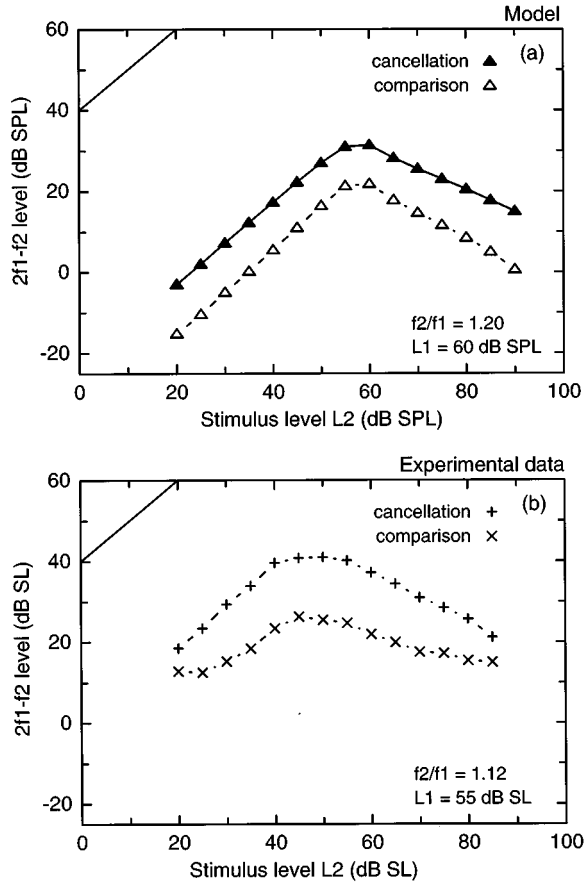


FIG. 6. Growth of the $2f_1-f_2$ combination tone with stimulus level L_2 for the case $L_1=$constant: (a) model results for $f_1=1400$ Hz and $f_2/f_1=1.20$ with $L_1=60$ dB SPL; (b) experimental results for one subject (GS) from Smoorenburg (1974) for $f_1=1000$ Hz and $f_2/f_1=1.12$ with $L_1=55$ dB SL.

perimental data. In the model, the gap increases monotonically from 2.3 to 10.8 dB for stimulus levels ranging from $L_1=L_2=20$ dB SPL to 80 dB SPL and $f_2/f_1=1.20$ [Fig. 5(a)]. This is similar to what one estimates by interpolating the psychoacoustic data for $f_2/f_1=1.12$ and $f_2/f_1=1.26$ [Fig. 5(b)]. Psychoacoustically, the difference between cancellation and comparison levels has been found to decrease with increasing frequency ratio f_2/f_1 of the primaries [Fig. 5(b); also Smoorenburg, 1974; Shannon and Houtgast, 1980]. It was verified that this is also the case in the model. For example, at $L_1=L_2=50$ dB SPL, the gap decreases from 10.9 dB to 3.8 dB for frequency ratios f_2/f_1 ranging from 1.125 to 1.33. These values are within the range of the available psychoacoustical data (Smoorenburg, 1972b, 1974; Shannon and Houtgast, 1980). The gap also decreases in the model with the order n of the CT frequency for fixed stimulus parameters. For example, at $L_1=L_2=50$ dB SPL and $f_2/f_1=1.20$, the gap decreases from 7.4 dB for the $2f_1-f_2$ CT ($n=1$) to 3.6 dB for the $3f_1-2f_2$ CT ($n=2$).

2. Case $L_1=$ constant

Figure 6(a) presents the growth of the $2f_1-f_2$ CT in the model with increasing L_2 for $L_1=60$ dB SPL and the same stimulus frequencies as in Fig. 5(a). The cancellation level increases proportionally with L_2 (slope 1.0 dB/dB) until a peak is reached near $L_2=55-60$ dB SPL. Thereafter, the cancellation level decreases with L_2 (slope -0.80 dB/dB). The growth of the comparison levels with L_2 shows a response pattern nearly identical to that of the cancellation levels. However, the comparison levels are on average 11.5 dB lower than the cancellation levels. The gap is essentially independent of L_2 from 20 to 80 dB SPL (range: 9.6–12.4 dB).

The model response curves are very similar to those found psychoacoustically, except for the somewhat lower cancellation and comparison levels obtained in the model for $L_2 \ll L_1$, and a peak that occurs at a slightly higher L_2 for equivalent f_2/f_1 and L_1 values. The available psychoacoustical data (Smoorenburg, 1972b, 1974; Shannon and Houtgast, 1980) also reveal a nearly constant gap between comparison and cancellation levels for $L_1=$constant. Figure 6(b) presents psychoacoustical data from Smoorenburg (1974) for the same subject as in Fig. 5(b) for $f_1=1000$ Hz and a frequency ratio $f_2/f_1=1.12$, the only ratio available. Limited accuracy of the pulsation-threshold technique near threshold might have affected the comparison levels found for low and high L_2 values.

III. TWO-TONE SUPPRESSION

A. Methodology

Under certain conditions, a stimulus component can be suppressed when presented simultaneously with a stronger tone or noise masker. This has been demonstrated both electrophysiologically (Sachs and Kiang, 1968) and psychoacoustically (Houtgast, 1972). Thus the probe in the cancellation-tone experiment could be subjected to suppression effects by the primary stimulus tones f_1 and f_2, and this could give rise to the difference between cancellation-tone (simultaneous) and comparison-tone (nonsimultaneous) methods of estimating the CT level.

Houtgast (1974) investigated several psychoacoustical techniques to quantify the amount of suppression by different maskers. It was found that the effect of suppression is revealed only when the masker and probe tones are presented nonsimultaneously, typically by using a forward masking, gap masking, or pulsation-threshold technique. Thus the suppression effects by the primaries during the cancellation experiment can be estimated using a variation of the comparison-tone method described in Sec. II C. Instead of primary tones f_1 and f_2, the stimulus now consists of a masker tone, typically f_1, and a tone at the CT frequency, typically $2f_1-f_2$, which is undergoing suppression. Psychoacoustically, the task is to match the loudness of the suppressed $2f_1-f_2$ tone with that of a probe tone at the same frequency but presented nonsimultaneously (Smoorenburg, 1974). In the model, this is reproduced by matching the BM excitation pattern produced by the probe tone with that of the stimulus component $2f_1-f_2$ during presentation with the masker f_1.

The model procedure is illustrated in Fig. 7 for the suppression of a stimulus tone $2f_1-f_2=1120$ Hz at a level of 31.4 dB SPL by a masker tone $f_1=1400$ Hz presented at $L_1=60$ dB SPL, in absence of $f_2=1680$ Hz. The masker f_1 and stimulus $2f_1-f_2$ tones are generated as cosine waveforms with zero phase. The thick line shows the BM excita-

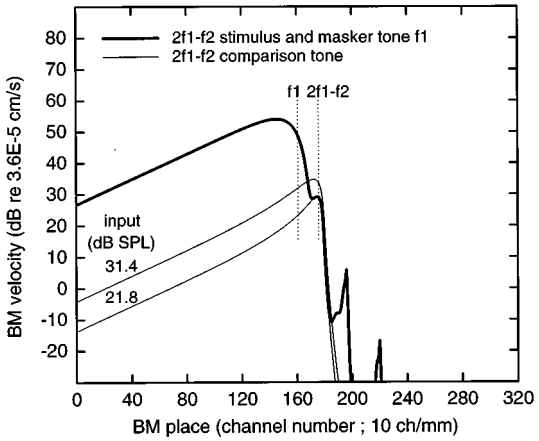


FIG. 7. BM excitation pattern for a stimulus tone $2f_1-f_2=1120$ Hz at 31.4 dB SPL suppressed by a masker tone $f_1=1400$ Hz presented at $L_1=60$ dB SPL (thick line), and for comparison tone $2f_1-f_2$ at two different levels (thin lines). The comparison level and phase of the suppressed tone are 21.8 db SPL and 29°, respectively. Note: dotted lines indicate the position of the low-level characteristic place corresponding to f_1 and $2f_2-f_1$.

tion pattern of this stimulus complex. The larger broad peak corresponds to masker f_1 and the second peak comes from the stimulus component $2f_1-f_2$. The interaction of these two stimulus components also generates combination tones which travel further along the cochlea. However, the focus in this type of experiments, both psychoacoustically and in the model, is to estimate the amount of suppression of stimulus component $2f_1-f_2$ by masker f_1. The two thin lines in Fig. 7 are the BM excitation patterns of a comparison probe tone at $2f_1-f_2=1120$ Hz, at two different input levels: 31.4 dB SPL, which is the original level of the $2f_1-f_2$ tone undergoing suppression, and 21.8 dB SPL, which is the level that provides the best match with the excitation pattern of the suppressed tone at the BM place for $2f_1-f_2$. As in the comparison-tone estimations of the CT in Sec. II C, the match is realized over the whole tip region of the excitation pattern at place $2f_1-f_2$. The model shows that, as a result of suppression, the peak of the tip at $2f_1-f_2$ has been reduced by an amount equivalent to 9.6 dB in stimulus level. The tip has also become narrower after suppression and shifted slightly toward the apex. These effects can be explained entirely in terms of the dependence of the place of maximum vibration and bandwidth of the BM peak upon level for single tones [Fig. 2(a)].

Additional simulations revealed that the phase of the traveling wave of stimulus component $2f_1-f_2$ is also affected by masker f_1. In the example of Fig. 7, the comparison probe tone at $2f_1-f_2$ must be generated with a phase lead of 29° with respect to component $2f_1-f_2$ in the stimulus complex to match both traveling wave phases at the BM place for $2f_1-f_2$. Thus in general, masker f_1 affects both the amplitude and phase of the stimulus component $2f_1-f_2$ undergoing suppression.

B. Results

Figure 8(a) illustrates the suppression of a stimulus component of frequency $2f_1-f_2=1120$ Hz by a masker tone f_1

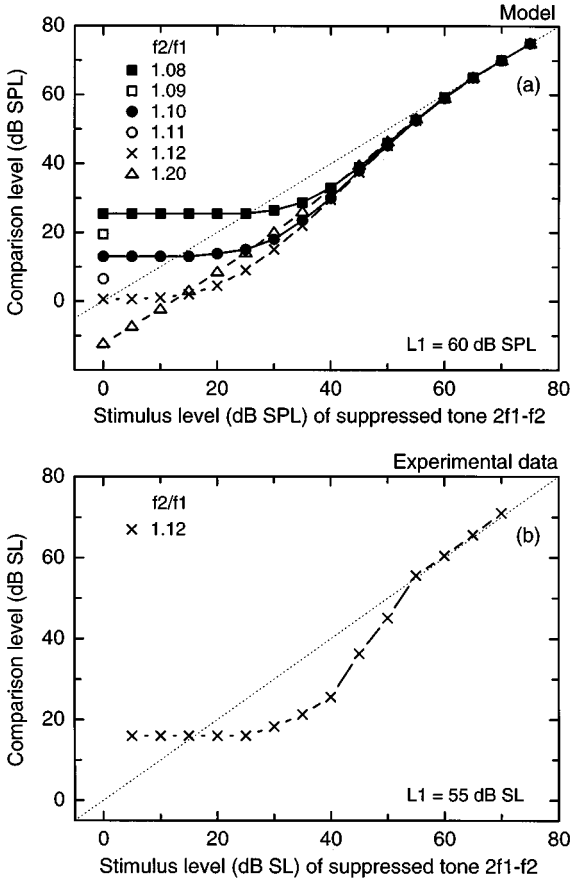


FIG. 8. Suppression of a tone of frequency $2f_1-f_2$ by f_1 only for constant L_1: (a) model results for $2f_1-f_2=1120$ Hz and $L=60$ dB SPL; (b) experimental results for one subject (GS) from Smoorenburg (1974) for $f_1=1000$ Hz and $L_1=55$ dB SL.

of level $L_1=60$ dB SPL in absence of f_2 for several f_2/f_1 ratios. The stimulus level of the $2f_1-f_2$ component is plotted on the abscissa, and the comparison level of the $2f_1-f_2$ probe tone is plotted on the ordinate. Data points along the dotted line indicates no suppression, i.e., the comparison level equals that of the stimulus component $2f_1-f_2$. This occurs in the model at all frequency ratios when the level of the stimulus component $2f_1-f_2$ exceeds the masker level L_1. For levels below L_1, increasingly smaller comparison-tone values than stimulus values are found. This is the suppression effect of masker tone f_1.

A plateau is reached at low levels when the activity introduced by the stimulus component $2f_1-f_2$ becomes smaller than that produced by the masker f_1 at the place corresponding to $2f_1-f_2$. In this case, the comparison level is obtained by matching the peak of the BM excitation pattern from the probe tone to the excitation pattern produced by f_1 at the place for $2f_1-f_2$. In Fig. 8(a), the height of the plateau increases with decreasing frequency ratio f_2/f_1, indicating increasingly greater masking by f_1 as would be expected. The height of the plateau is very sensitive to small variations in f_2/f_1. This is due to the very steep slope of the BM excitation pattern produced by f_1 on the apical side.

The shapes of the suppression curves shown by the model closely resemble those obtained psychoacoustically.

Figure 8(b) presents pulsation-threshold data from Smoorenburg (1974) for the same subject as in Figs. 5(b) and 6(b) for the suppression of a stimulus component $2f_1-f_2$ by a masker tone $f_1=1000$ Hz of level $L_1=55$ dB SL in absence of f_2 and at a ratio $f_2/f_1=1.12$. Again, there is suppression only when the stimulus level of the $2f_1-f_2$ component is below the level L_1 of the masker, and there is a plateau at low levels. The height of the plateau in Fig. 8(b) is higher than in Fig. 8(a) for the same f_2/f_1, probably an indication that the apical slope of the excitation pattern produced by f_1 is steeper in the model than for the subject.

IV. DISCUSSION

The model and psychoacoustical data presented in Figs. 5 and 6 clearly indicate a significant difference between cancellation-tone and comparison-tone methods of estimating the CT level. The level L_1 of the stimulus complex seems to be the key parameter controlling the difference between cancellation and comparison values. Essentially, for a fixed frequency ratio f_2/f_1, the level L_1 largely determines the gap between cancellation and comparison levels. The working hypothesis is that the higher levels found for simultaneous probe tones, as in the cancellation-tone experiment, can be attributed to suppression effects by the primary tone f_1 (Smoorenburg, 1972b, 1974). In other words, the cancellation tone could be suppressed by f_1 as it travels along the cochlea to the point of cancellation, and thus a higher level is necessary to counteract the internally generated CT.

The two-tone suppression data presented in Fig. 8 allow testing of the above hypothesis by quantifying the suppression of the cancellation tone $2f_1-f_2$ by the primary tone f_1. If the hypothesis is true, then the comparison-tone estimates of the CT level could be predicted from the cancellation-tone estimates by subtracting from the latter the amount of suppression produced by f_1 on the cancellation tone. By and large, quantitative comparisons confirm the hypothesis. For example, the model comparison levels of the $2f_1-f_2$ CT in Fig. 6(a) ($f_1=1400$ Hz, $L_1=60$ dB SPL, and $f_2/f_1=1.20$) are predicted within 0.25 dB for $L_2=20$ dB SPL to 80 dB SPL from the model cancellation levels in the same figure and the two-tone suppression data from Fig. 8(a). Above $L_2=80$ dB SPL, this prediction procedure slightly overestimates the model comparison levels, which indicates an underestimation of the suppression effects on the cancellation tone. At $L_2=90$ dB SPL the prediction error reaches 2.4 dB. Further analyses with the model revealed that, above $L_2=80$ dB SPL, the BM excitation pattern produced by f_2 is no longer negligible at the place corresponding to $2f_2-f_1$ compared to that produced by f_1, whose level is fixed at 60 dB SPL. Thus the small error between model comparison levels and predicted values when $L_2 \gg L_1$ is likely due to suppression of the cancellation tone by f_2 in addition to that arising from f_1.

The prediction procedure detailed above can also be tested with the psychoacoustical data presented in Figs. 6(b) and 8(b). This is the only complete set of comparison-tone, cancellation-tone, and two-tone suppression data obtained with the same stimulus conditions ($f_1=1000$ Hz, $L_1=55$ dB SPL, and $f_2/f_1=1.12$) for that subject. Again, the difference between cancellation-tone and comparison-tone estimates of the CT in Fig. 6(b), which reaches up to 15.5 dB, can largely be explained by the two-tone suppression data from Fig. 8(b). The error in predicting the psychoacoustical comparison levels of the CT from the cancellation levels and the two-tone suppression data has a mean of 1.7 dB over the entire set of L_2 values. It always corresponds to an overestimation of the comparison levels of the CT (range: $+0.4$ to $+3.5$ dB). However, given that the prediction procedure involves the subtraction of two psychoacoustic measurements (cancellation levels for the CT and comparison levels for two-tone suppression) and the comparison to a third one (comparison levels for the CT), the prediction error should be well within the total measurement error. Shannon and Houtgast (1980) applied a similar procedure on their psychoacoustical data for a range of stimulus conditions, and found a good agreement in three out of four subjects.

The model and psychoacoustical predictions just described support the hypothesis that the cancellation tone is suppressed by f_1, and that this is the main reason for the difference between cancellation and comparison levels of the CT. However, they do not address the effect of suppression on the phase of the traveling wave corresponding to the cancellation tone. The model simulations presented in Figs. 4 and 7 provide some insight into this question. In Fig. 4, the cancellation-tone (31.4 dB SPL) and comparison-tone (21.8 dB SPL) estimates of the $2f_1-f_2$ CT level differ by 9.6 dB. Moreover, the cancellation phase ($-17°$) is not exactly 180° out-of-phase with the comparison phase (192°); the mismatch amounts to a comparison phase lead of 29°. Figure 7 shows that under the same stimulus conditions f_1 and L_1, but in absence of f_2, a stimulus component $2f_1-f_2$ presented at 31.4 dB SPL is suppressed by f_1 to the level of a 21.8 dB SPL tone. Moreover, its traveling wave undergoes a phase lead shift of 29° near the BM place for $2f_1-f_2$ (see Sec. III A). Thus when the suppression effects of f_1 are accounted for in the cancellation-tone experiment, the traveling wave of the cancellation tone has the same amplitude and is in exact out-of-phase opposition with the internally generated CT near the BM place for $2f_1-f_2$. These two conditions are necessary for destructive interference. The exact amount of cancellation-tone amplitude suppression and phase shift due to f_1 will, however, vary according to the stimulus conditions (mainly f_2/f_1 and L_1).

The suppression effect of f_1 on the traveling phase of the cancellation tone found in the model simulations has a direct implication for the psychoacoustic pulsation-threshold technique. For CT measurements, the phase opposition assumption between cancellation and comparison tones used to derive a stimulus phase value for the comparison tone (Smoorenburg, 1972b) may not be valid. For two-tone suppression measurements, a phase shift of the comparison tone may be required. In practice, however, the gradual mixing of the CT during stimulus presentation and the alternating probe tone from complimentary on/off gates (typically 15 ms) may well result in an audible phase modulation with little effect on the pulsation threshold.

V. CONCLUDING REMARKS

The model results presented in this study and the reviewed psychoacoustical data indicated significantly higher estimates of the $2f_1-f_2$ combination-tone level using a cancellation-tone method than a comparison-tone method. In the former the stimulus and the probe tone are presented simultaneously, in the latter they are presented nonsimultaneously. The gap between the two methods increases with the stimulus level $L_1=L_2$ of the two primary tones f_1 and f_2. The gap is also much more sensitive to a change in L_1 than L_2. It also decreases with increasing frequency ratio f_2/f_1 and increasing order n of the combination tone $f_1-n(f_2-f_1)$, where $f_2>f_1$. Hence, the critical parameters controlling the difference are the level L_1, and the frequency separation between f_1 and the combination tone to be measured.

This study also strongly supports the idea that the difference between cancellation and comparison levels is mainly due to suppression effects by f_1 (Smoorenburg, 1974; Shannon and Houtgast, 1980), and thus that psychoacoustic combination-tone generation and two-tone suppression phenomena have a common origin in the nonlinearity of the cochlea. On the basis of the model results, the cancellation-tone experiment could be interpreted as follows. Cancellation occurs on the BM when the traveling wave of the f_1 suppressed external probe tone at $2f_1-f_2$ has the same amplitude and becomes in exact phase opposition with the internally generated $2f_1-f_2$ distortion component. There is also some evidence of suppression effects by f_2 in the model when $L_2 \gg L_1$. The cancellation amplitude and phase conditions are met over a substantial region of the BM near the characteristic place for $2f_1-f_2$.

A major simplification in the model is the lack of a second source of filtering (Duifhuis, 1976). This could be implemented, for example, by using a formal micromechanics model of the tectorial-basilar membrane interaction (Allen and Fahey, 1993), or by including a stage of frequency selectivity within the nonlinear pressure source (Kanis and de Boer, 1993). This would make it possible to use larger feedback gains for the nonlinearity, in particular the case for which the BM damping becomes negative. A related question is whether the nonlinearity should be applied to the BM damping, as in this study, or to another mechanical element, like BM stiffness. In addition to providing potentially higher and more realistic cochlear wave amplification, these model improvements could be particularly important to achieve a good reproduction of the cancellation phase and of the sensitivity of the CT with the frequency ratio f_2/f_1 of the primaries. However, our main conclusion regarding the importance of the suppression effects by f_1 to explain the difference between cancellation and comparison level estimates of the CT within a given model or subject would be retained.

Finally, there is a great deal of interest currently in the study of DPOAEs from ear canal measurements (Probst *et al.*, 1991). This stems from the objective nature of the measurements, the availability of commercial instruments, and the clinical diagnostic potential of this approach to probe the cochlear function, among other reasons. However, the psychoacoustic approach of investigating auditory distortion products and suppression effects should not be overlooked. First, it does not require to make any assumption about the retrograde propagation through the cochlea and middle ear. Thus psychoacoustic combination-tone and suppression data provide a very direct link to probe the important active and nonlinear parameters of the cochlea, and together they impose stringent requirements on models. Moreover, the psychoacoustic data may be inherently less variable than the DPOAE data from subject to subject on the basis that the external cancellation or comparison tone used to collect the psychoacoustic data passes through the same system as the distortion product to be measured. The benefits of this type of normalization may merit further attention.

ACKNOWLEDGMENTS

This work was conducted as part of a postdoctoral research fellowship awarded to the first author by the Natural Sciences and Engineering Research Council of Canada. Additional financial support was provided by the Helmholtz Research School for Autonomous Systems Research.

Allen, J. B., and Fahey, P. F. (**1993**). ''A second cochlear-frequency map that correlates distortion product and neural tuning measurements,'' J. Acoust. Soc. Am. **94**, 809–816.

Duifhuis, H. (**1976**). ''Cochlear nonlinearity and second filter: Possible mechanism and implications,'' J. Acoust. Soc. Am. **59**, 408–423.

Fettweis, A. (**1986**). ''Wave digital filters: Theory and practice,'' Proc. IEEE **74**, 270–327.

Fukusawa, T., and Tanaka, Y. (**1994**). ''Distortion product otoacoustic emissions in an active nonlinear model of the cochlea,'' Hearing Res. **81**, 42–48.

Giguère, C., and Woodland, P. C. (**1994**). ''A computational model of the auditory periphery for speech and hearing research. I. Ascending path,'' J. Acoust. Soc. Am. **95**, 331–342.

Giguère, C., Kunov, H., and Smoorenburg, G. (**1995a**). ''Computational modeling of psychoacoustical combination tones and distortion-product otoacoustic emissions,'' in Proceedings of the 15th Int. Cong. on Acoustics, Trondheim, Norway, Vol. III, pp. 237–240.

Giguère, C., Kunov, H., and Madsen, P. (**1995b**). ''Distortion-product otoacoustic emissions: computer modeling and experimental data,'' Can. Acoust. **23**, 25–26.

Goldstein, J. L. (**1967**). ''Auditory nonlinearity,'' J. Acoust. Soc. Am. **41**, 676–689.

Goldstein, J. L. (**1970**). ''Aural combination tones,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden), pp. 436–444.

Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Helle, R. (**1969**). ''Amplitude und Phase des im Gehörs gebildeten Differenztones dritter Ordnung,'' Acustica **22**, 74–87.

Houtgast, T. (**1972**). ''Psychophysical evidence for lateral inhibition in hearing,'' J. Acoust. Soc. Am. **51**, 1885–1894.

Houtgast, T. (**1974**). ''Lateral suppression in hearing: a psychophysical study on the ear's capability to preserve and enhance spectral contrasts,'' Ph.D. dissertation, Academische Pers B.V., Amsterdam.

Humes, L. E. (**1983**). ''Psychophysical measures of two-tone suppression and distortion products ($2f_1-f_2$) and (f_2-f_1),'' J. Acoust. Soc. Am. **73**, 930–950.

Johnstone, B. M., Patuzzi, R., and Yates, G. K. (**1986**). ''Basilar membrane measurements and the traveling wave,'' Hearing Res. **22**, 147–153.

Kanis, J. L., and de Boer, E. (**1993**). ''Self-suppression in a locally active nonlinear model of the cochlea,'' J. Acoust. Soc. Am. **94**, 3199–3206.

Kemp, D. T. (**1979**). ''Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea,'' Arch. Otorhinolaryngol. **224**, 37–45.

Lutman, M. E., and Martin, A. M. (**1979**). ''Development of an electroacoustic analogue model of the middle ear and acoustic reflex,'' J. Sound Vib. **64**, 133–157.

Patuzzi, R. B., Yates, G. K., and Johnstone, B. M. (**1989**). ''Outer hair cell receptor current and sensorineural hearing loss,'' Hearing Res. **42**, 47–52.
Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). ''A review of otoacoustic emissions,'' J. Acoust. Soc. Am. **89**, 2027–2067.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). ''Mössbauer measurements of the mechanical response to single-tone and two-tone stimuli at the base of the chinchilla cochlea,'' in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, New York), pp. 121–128.
Ruggero, M. A. (**1992**). ''Responses to sound of the basilar membrane of the mammalian cochlea,'' Curr. Opin. Neurobiol. **2**, 449–456.
Sachs, M. B., and Kiang, N. Y. S. (**1968**). ''Two-tone inhibition in the auditory nerve,'' J. Acoust. Soc. Am. **43**, 1120–1128.
Shannon, R. V., and Houtgast, T. (**1980**). ''Psychophysical measurements relating suppression and combination tones,'' J. Acoust. Soc. Am. **68**, 825–829.
Smoorenburg, G. F. (**1972a**). ''Auditory region of combination tones,'' J. Acoust. Soc. Am. **52**, 603–614.
Smoorenburg, G. F. (**1972b**). ''Combination tones and their origin,'' J. Acoust. Soc. Am. **52**, 615–632.
Smoorenburg, G. F. (**1974**). ''On the mechanisms of combination tone generation and lateral inhibition in hearing,'' in *Psychological Models and Physiological Facts in Hearing*, edited by E. Zwicker and E. Terhardt (Springler-Verlag, Berlin), pp. 332–343.
Smoorenburg, G. F., Gibson, M. M., Kitzes, L. M., Rose, J. E., and Hind, J. E. (**1976**). ''Correlates of combination tones observed in the response of neurons in the anteroventral cochlear nucleus of the cat,'' J. Acoust. Soc. Am. **59**, 945–962.
Strube, H. W. (**1986**). ''The shape of the nonlinearity generating the combination tone $2f_1-f_2$,'' J. Acoust. Soc. Am. **79**, 1511–1518.
Sun, X.-M., Schmiedt, R. A., He, N.-J., and Lam, C. F. (**1994a**). ''Modeling the fine structure of the $2f_1-f_2$ acoustic distortion product. I. Model development,'' J. Acoust. Soc. Am. **96**, 2166–2174.
Sun, X.-M., Schmiedt, R. A., He, N.-J., and Lam, C. F. (**1994b**). ''Modeling the fine structure of the $2f_1-f_2$ acoustic distortion product. I. Model evaluation,'' J. Acoust. Soc. Am. **96**, 2175–2183.
Zwicker, E. (**1955**). ''Der ungewöhnliche Amplitudengang der nichtlinearen Verzerrungen des Ohres,'' Acustica **5**, 67–74.
Zwicker, E. (**1968**). ''Der kubische Differenzton und die Erregung des Gehörs,'' Acustica **20**, 206–209.

Spontaneous otoacoustic emission frequency is modulated by heartbeat

Glenis R. Long
Department of Audiology and Speech Sciences, Purdue University, West Lafayette, Indiana 47907

Carrick L. Talmadge
Department of Physics, Purdue University, West Lafayette, Indiana 47907

(Received 21 April 1997; accepted for publication 8 August 1997)

Detailed analysis of spontaneous otoacoustic emissions (SOAEs) in human subjects revealed that all stable SOAEs sufficiently above the noise floor to permit appropriate analysis have sidebands at multiples of approximately 1 Hz. This is consistent with the hypothesis that SOAEs are modulated by heartbeat. Simultaneous measurement of the rate of blood flow to the thumb and the separation of the spectral sidebands demonstrated that they covary ($r=0.982$, $p<5\times10^{-10}$). An adaptive least-squares fit (LSF) paradigm was developed to facilitate the measurement of the instantaneous frequency and amplitude of the signals. A combination of traditional spectral analyses and new LSF analyses showed that the sideband generation stems from frequency modulation of the emissions. If there is any amplitude modulation correlated with the blood flow, it is below the noise floor of the analysis. The frequency of the emission was at a minimum when the blood flow was maximal. Examination of alternative mechanisms using computer simulations suggests that these changes stem from changes of 10–100 ppm in the mass of the basilar membrane. © *1997 Acoustical Society of America.* [S0001-4966(97)06311-X]

PACS numbers: 43.64.Bt, 43.64.Jb, 43.64.Yp [BLM]

INTRODUCTION

The discovery that the healthy cochlea generates sounds (Kemp, 1978), known as otoacoustic emissions, has revolutionized the understanding of cochlear function (for review see Dallos, 1996; Patuzzi, 1996). These sounds can be generated either in the absence of acoustic stimulation (spontaneous otoacoustic emissions) or in response to acoustic stimulation (evoked otoacoustic emissions).

Spontaneous otoacoustic emissions (SOAEs) are noise-perturbed sinusoidal acoustic signals (e.g., Talmadge *et al.*, 1991). The analysis of signals recorded from a sensitive microphone placed at the entrance to the ear canal provides a noninvasive tool for monitoring the state of the cochlea. Spontaneous emissions are indicators of the active nonlinear processes responsible for the remarkable sensitivity of the auditory system. Evidence for this claim is provided by their vulnerability to anything that impairs the sensitive nonlinear component of cochlear mechanics. For example, the levels and/or frequencies of spontaneous emissions can be changed by aspirin (for review see Probst *et al.*, 1991; Long *et al.*, 1991), quinine (for review see McFadden and Pasanen, 1994), and contralateral stimulation (Mott *et al.*, 1989), all of which reduce the sensitive nonlinear component of basilar membrane vibration (Murugasu and Russell, 1996; Russell and Murugasu, 1996). As a result, they have the potential to serve as very sensitive indicators of the physiological condition of the cochlea.

The outer ear is a very noisy environment and SOAEs are low level signals (the majority are below 0 dB SPL). Consequently, it is necessary to use signal processing to examine them. The most common analysis is in the spectral domain using discrete Fourier transforms (see for examples, Talmadge *et al.*, 1993a). When the discrete Fourier transform used has permitted fine-grained analysis, reproducible sidebands at frequencies consistent with heartbeat modulation (approximately 1 Hz arising from the ~60 beats/min=1 beats/s) have been observed in adult human subjects (Talmadge *et al.*, 1991; Rao and Bilger, 1991; Bell, 1992; van Dijk *et al.*, 1994). When time domain analysis has been used, frequency modulations of 1 Hz were observed in SOAEs from adults (van Dijk *et al.*, 1994) and an approximately 2 Hz modulation was detected in SOAEs from normal infants who have a heartbeat of about ~120 beats/min (Burns and Pitton, 1993). SOAEs found in 5 out of 126 young pigmented guinea pigs had sidebands at approximately 5 Hz, consistent with the heartbeat rate of this species of approximately ~300 beats/min (Ren *et al.*, 1995). These authors did not determine if the signals were frequency modulated, but found evidence for amplitude modulation of the signal by time averaging the output of a 50-Hz bandwidth filter centered around the frequency of the SOAE.

Normal blood flow to the cochlea is important for the maintenance of normal cochlear function because cochlear vessels provide the metabolites for the maintenance of cochlear tissues and fluids. Impaired cochlear blood flow has been hypothesized to be the basis of a variety of inner ear disorders, including Ménière's disease, noise damage, and sudden hearing loss (Axelsson and Ryan, 1988). Consequently, blood flow based modulation of SOAEs by heartbeat could provide a useful tool for evaluation of the condition of the cochlear microcirculation and thus provide useful information for the diagnosis and treatment of inner ear disorders. Aspects of spontaneous firing of auditory nerve fibers and reaction time to tones which depend on the cardiac cycle (reviewed in Ren *et al.*, 1995) provide additional evidence

that cochlear function is modulated by heartbeat.

This paper presents evidence that (a) the sidebands are a regular feature of spontaneous otoacoustic emissions in humans when the analysis is optimized for their detection, (b) the frequency separations between sidebands is highly correlated with heartbeat frequency, and (c) the sidebands in human subjects are generated dominately by frequency modulation of the emissions and not by amplitude modulation. Approximate analytic results based upon the limit-cycle oscillator formalism (Talmadge *et al.*, 1991) were used to evaluate possible mechanisms responsible for the observed frequency modulation. Computer simulations of the same model were also used to verify the accuracy of the analytic results.

I. GENERAL METHODS

A. Subjects

Some of the subjects were a subset of the subjects presented in Talmadge *et al.* (1993a). Other subjects were not included in that study because they represented biased population points (they had previously been screened for emissions). Measurements were obtained in two laboratories at Purdue, one in the Department of Audiology and Speech Sciences (AUS) and the other in the Physics Department.

B. Apparatus

Subjects were seated comfortably in a double-walled IAC sound-treated booth. An Etymotic Research ER-10 low-noise microphone was inserted in the ear canal. The ear canal signal was bandpass filtered and amplified (300–10 000 Hz, roll off=6 dB/octave, gain=200) using a Stanford SR560 low-noise amplifier. A twin-conductor shielded cable (balanced microphone cable) was used to electrically connect the amplifier to the digitization and recording equipment. Contrary to the usual procedure in our laboratory, recordings were taken almost immediately after the subject had entered the booth without allowing 10–15 min for the emissions to stabilize.

C. Data collection

1. SOAE recordings

Five-minute (300-s) recordings were obtained from each ear using the ER10 microphone and the associated preamplifier. The signals were filtered by the low-noise amplifier and then digitized in 16-bit samples at a 44 100-Hz clock rate directly to disk using a Singular Solutions A/D 64× analog-to-digital converter digitally connected to a NeXT computer. The signals were later analyzed offline. To conserve disk space, and also to speed up the analysis, the samples were later converted to a lower sampling rate (''down sampled'') using a time-domain digital low-pass filter, which was applied using the NeXT 56001 DSP coprocessor. Typically the down-sampling factor was 3, which gave a reduced sampling rate of 14 700 Hz.

The digital filter was a 201-point finite-impulse-response filter, with the size being determined by the available amount of DSP memory in the NeXT computer. The filter coefficients were generated by taking the product of a sinc function $[\sin(x)/x]$ and a Hamming window. In particular, given a sequence of samples, y_n $(n=0,...,N_y-1$, where N_y is the total number of recorded samples), then the filtered signal $\tilde{y}_l$ is given by

$$\tilde{y}_l=\sum_{k=-N_w}^{N_w} W_k\cdot Y_{k+l}, \tag{1}$$

where $N_w=(201-1)/2=100$. Here W_k $(k=-N_w,\ldots,+N_w)$ are the filter coefficients given by

$$W_k=N_0\,\frac{\sin\,\pi k f_c\Delta t}{\pi k f_c\Delta t}\cdot H_k, \tag{2}$$

$$H_k=\left[0.54+0.46\cos\frac{\pi k}{N_w}\right], \tag{3}$$

where $\Delta t=1/(44\,100\ \text{Hz})$ is the digital sampling rate, f_c is the cutoff frequency of the filter, and N_0 is chosen so that $\Sigma_k W_k=1$. The first and last 100 points of the filtered signal were discarded to remove end effects. After the digital filter was applied, the recorded samples were decimated in time to reduce the total sample size, which for a down-sampling factor of three corresponded to a new sample size of $N_y/3$.

Verification that the digital filter did not produce artifacts in the down-sampled signal was made by comparing spectra of the signal before and after down-sampling. An example of such a spectrum is given in Fig. 1. It was found in these studies that in the frequency range of interest (<6500 Hz), the spectral amplitudes of the pre- and post-down-sampling periodogram were nearly identical. The existence of a possible time delay from the digital filter was also studied by comparing the waveforms before and after filtering. The delay induced by the filter was approximately -3 ms, which is in agreement with the number of points (100) discarded from the beginning of the filtered data.

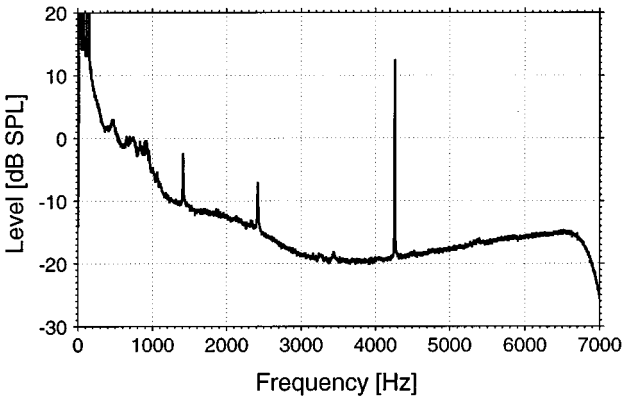


FIG. 1. Spectral periodogram of a 386-s recording for subject LN (run 941219R1). The spectrum was obtained using 1.1-s window size via the technique described in the text.

2. Blood flow recordings

In studies 2 and 3, blood flow measurements were also made by placing a photo-plethysometer over the right thumb nail of each experimental subject. A velcro strap held the assembly in place. The signal from the photo-plethysometer

was amplified using an Acudata Model 109 preamplifier and subsequently low-pass filtered at 30 Hz using a Wavetek 753A brick-wall filter in order to remove 60-Hz contamination from the signal (introducing a delay of approximately +27 ms). This signal was also digitized at a 44 100-Hz clock rate using the analog-to-digital converter, and was later down-sampled to 490 Hz using a 100-point digital low-pass filter. This filter was generated using Eqs. (1)–(3) as before and had a delay of −100 ms for a 1-Hz signal.

Since the blood flow measurements were taken at the fingertip of the subject, this raises the question as to whether a significant phase difference between the blood flow at the fingertip and at the cochlea could be present. This question can be addressed by considering measurements of the pulse wave velocity. The velocity of the pulse wave depends on the distensibility of the arterial wall and thus increases with age and is faster in the periphery. In a 25 year old, the pulse wave velocity is greater than 6 m/s in the aorta, and increases by 80% in the leg arteries (Holtz, 1996). The difference in distance of the fingertip and the cochlea from the heart stems from the division of the brachiocephalic trunk into the carotid and subclavian arteries. The distance was estimated to be less than 50 cm for the subjects used in this study. Since both the carotid and subclavian articles are much narrower than the aorta, the pulse wave velocity will be greater than 10 m/s. It is concluded from this analysis that the combined effect of the delay from the Wavetek filter, the digital filters, and the difference in timing of blood flow at the finger tip and the inner ear should be less than 60 ms in normal human subjects. This time delay sets a limit on the corresponding phase error of 22° for a 1-Hz signal, which is an acceptable level of error for this study.

D. Data analysis

1. Spectral periodogram

All traditional spectral measurements (e.g., Fig. 1) reported in this paper were performed using the spectral periodogram methodology described in Talmadge *et al.* (1993a). This technique involves averaging the amplitudes of fast Fourier transforms (FFTs) obtained from successive three-quarter overlapping time windows. In all cases, a Welch window function (Press *et al.*, 1992) was applied to each time-domain window before the FFT was obtained. The peak excursion of the signal in each time window was also compared against a preset threshold level, which was used to reject noisy data in obtaining the spectral periodogram.

2. Centered spectral averaging

Spontaneous emissions are not stable in frequency, but tend to drift about some central frequency over time, and especially near the beginning of the session (Zurek, 1981; Whitehead, 1991). This frequency drift (seen in Fig. 2) leads to a ''smearing out'' of any fine structure that might be present in the emission spectrum. For instance, the gray line in Fig. 3 displays the typical shape of a spectral line using a conventional spectral periodogram. While this spectrum does not correspond to the classical Lorentzian line shape that might be expected for a noise-driven oscillator, it also shows no clear evidence of any frequency splitting of the central frequency.

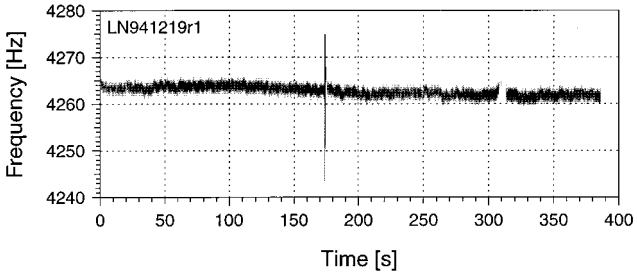


FIG. 2. Sonogram of the 4264-Hz SOAE for subject LN. Frequency is plotted as a function of time, with a darker shades corresponding to more intense sound levels. The discontinuities at 172 s and 310 s are due to acoustical artifacts (e.g., coughing).

The centered-averaging method entails computing a series of (131 072 samples=11.9 s) spectral periodograms from the full recording (of typical length 300 s and 44 100-Hz sampling rate), and shifting the central peak of each spectrogram for a particular SOAE to the average frequency of the central peak for the *entire* data sample. The 131 072-sample spectral periodograms were obtained by summing the amplitudes of FFTs from successive 3/4 overlapping short (65 536 samples=4.45 s) data segments. To improve the estimation of the central frequency of the spectral periodogram, the FFTs were zero padded by a factor of 4 (raising the number of points per FFT to 65 536 samples×4 =262 144 points) to give a factor of 4 improvement in estimation of the central frequency of the SOAE for the resulting FFT. Finally, the power spectrograms obtained in this way were summed together to produce the centered average power spectral periodogram. This procedure substantially reduced the long-term frequency drift normally present in the recordings of SOAEs. As with the traditional spectral periodogram, threshold rejection based on peak excursion was used to remove noisy data segments from the averaged spectrum. The black line in Fig. 3 displays a typical SOAE spec-

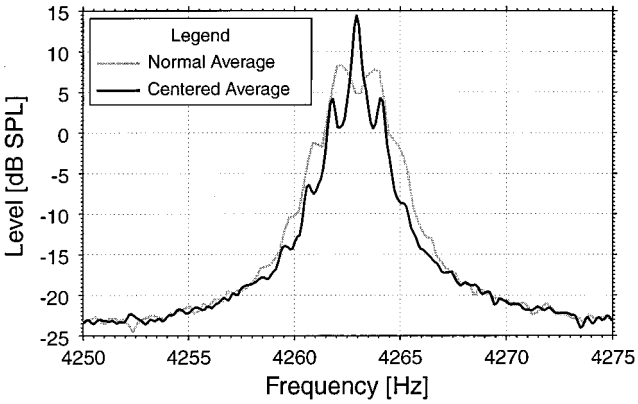


FIG. 3. Comparison of spectral line shapes for subject LN (run 941219R1), using the conventional ''normal'' (gray line) and the centered (black line) spectral averaging techniques. The spectra are obtained from a 386-s recording for subject LN.

tral line shape obtained using centered averaging.

3. Measuring SOAE amplitude and frequency

Although the FFT can be used to obtain short length spectra (e.g., using 0.5-s intervals), from which the time-varying SOAE peak frequencies and levels can be determined, this significantly limits the frequency resolution (in this example, to 2 Hz). The frequency resolution can be increased by increasing the time window length. However, since a 1-Hz modulation rate (arising from the heartbeat rate in humans) of either SOAE frequency or amplitude modulation is being searched for, any modulation would be significantly ''smoothed out,'' were the window size to be made much larger than 0.5 s.

When using FFTs to obtain SOAE frequency as a function of time in this study, 8192-sample Welch-windowed FFTs were calculated, with an offset of 147-points between consecutive windows. This choice of offset gave an effective 0.01-s separation between consecutive frequency determinations (the sampling rate of the data file was 14 700 Hz). The resulting FFTs were used to estimate the time rate of change of the amplitudes and frequencies of the SOAEs of a given signal recording. The overlap choice resulted in SOAE amplitude and frequency ''signals'' digitized at a sampling rate of $1/(0.01\ \mathrm{s}) = 100$ Hz. This rate was chosen so that the resulting amplitude and frequency fluctuations could themselves be analyzed via an FFT: The 100-Hz sampling rate corresponds to a 50-Hz Nyquist limit, which is safely larger than any observable harmonic of the 1-Hz heartbeat rate.

The 8192-sample windows were zero padded to 131 072 points (zero-padding factor of 16) to improve their resolution. However, even this very large window size results in only a ~ 0.112-Hz resolution, which is not adequate. Further resolution improvements were made by fitting the log amplitude of the SOAE about its maximum value to a second-order (quadratic) polynomial. All neighboring points with an amplitude within 12 dB of the peak amplitude were used in the polynomial fit, corresponding typically to 12–14 adjacent frequency bins being used in the obtaining the best fit of the polynomial. The entire procedure was extremely computationally expensive. Analyzing a single SOAE from a 386-s recording (LN941219r1) took more than 12 h on a Hewlett-Packard 712/80 system (floating point computations benchmark at over 30 MFLOPs on this system).

It should be noted that quadratic peak fitting has been criticized as inaccurate by Brown and Puckette (1993). However, this criticism was based on performing 256-point nonzero-padded FFTs, and using only three adjacent frequency bins to obtain the quadratic-polynomial fit. In contrast, the current method uses 8192-point data segments with a zero-padding factor of 16. Consequently a much better estimate of the central peak can be obtained, since 12–14 points are used in this method for the quadratic-polynomial fit. The differences between two analysis methods can be explained in terms of differences in the goals of the two studies. Brown and Puckette intended that the analysis be performable in real time, and a factor that was not a consideration in the current study. The quadratic peak fitting method as used in the current study was found to be of comparable precision and accuracy to the other methods tested, and was not used only because of its prohibitive computational expense.

4. Least-squares fit time-domain filtering

Because the quadratic peak fitting method is computationally expensive to perform, an alternative technique known as least-squares fit (LSF) filtering was employed to perform the bulk of the calculations. The quadratic peak fitting method was used in a few cases to ''spot check'' the accuracy of the LSF filtering method. As detailed in the Appendix a, the LSF algorithm consists of assuming a particular model for the signal that is embedded in a noise background, and then determining the parameters of that model which give the least-squares fit of the model to the data. In this case it was assumed that the signal is a pseudo-stable sinusoidal signal with a previously (but inaccurately) determined frequency. The purpose of the LSF technique is to extract both the amplitude and frequency of the SOAE as a function of time, where the SOAE is assumed to oscillate near the preset frequency level.

The LSF method allows one to determine the frequency of a single signal component in a given window size to a much higher precision than is possible with a comparable FFT window size. For instance, the FFT takes a window size of at least 1 s in length to determine a frequency to within 1 Hz. The LSF method, in contrast, is able to obtain an estimate for both the level and frequency of the signal component in a small fraction of a second. The length of time needed to resolve this component depends on both the sampling rate as well as sample length and the signal-to-noise ratio in the system. This dependence differs from the FFT algorithm, where increasing the sampling rate only increases the maximum resolvable frequency and does not improve the resolution of intermediate frequencies. The LSF algorithm also builds in correlations between all of the frequency components being analyzed, allowing this method to resolve adjacent frequency components that cannot ordinarily be resolved using simple FFTs.

The difference in speed between the two techniques can be understood partly in terms of the observation that only the frequencies being investigated need be analyzed using the LSF technique, whereas using the FFT algorithm the full range of frequency components (0 Hz to the Nyquist limit) must be examined in equally spaced frequency intervals at the desired frequency resolution. The FFT algorithm allows one to analyze N frequency components in a time proportional to $N \log(N)$. However, the LSF technique, as outlined in the Appendix a, is of order $m \cdot N$, where m is the number of frequency components under consideration, and N is the number of points in the time window for the LSF filter. Thus the FFT is of order $\log(N)/m$ times slower than the LSF filter. For the recording mentioned above (LN941219r1), the total processing time for the LSF technique was approximately 360 s.

When analyzing SOAEs, the window length of the LSF filter was typically chosen to be near 0.5 s. For those data samples which were down sampled to 14 700 Hz, this corresponded to a window size of 8192 points, and a window

length of ~0.56 s. (The window size was chosen to be a power of two only for the convenience of comparison with FFT-based techniques, since there are no similar restrictions on the window size for the LSF method.) Once again a window offset of approximately 0.01 s was chosen, since this resulted in a sampling rate of SOAE amplitude and frequency at 100 Hz, allowing these quantities to be further analyzed for rate and level of modulation.

5. Adaptive least-squares fit time-domain filtering

The Appendix provides an algorithm for constructing a LSF filter to estimate the frequency of an emission given an an initial estimate of its frequency. However, this LSF filter has the potential disadvantage that the center frequency of the filter is fixed to the initial estimate of the emission frequency. This algorithm was improved upon for some analyses by allowing the program to iteratively determine the emission frequency for a particular time window. An initial estimated value of the emission frequency was supplied to the LSF software routine, which used it to obtain a more exact frequency using the LSF fit. This improved frequency estimate was fed back into the LSF filter to further refine the estimated emission frequency. The feedback procedure was repeated until either the emission frequencies returned from the LSF filter stabilized about a particular value to within a preset precision, or a preset maximum number of iterations had been performed. When the iterative LSF filter technique succeeded in meeting the precision goal, this new frequency estimate was used as the initial estimate for the emission frequency in the following time window. In the rare instances when the frequency estimate did not stabilize within the maximum allowed number of iterations, the estimate of the emission frequency at the beginning of the session was used to restart the iterative procedure in the following time window.

The iterative technique has the advantage that it allows the central frequency of the LSF filter to vary over time in order to track the instantaneous frequency of the SOAE. The technique of allowing the filter parameters for a given time window to depend on the emission frequency from the previous time window in this study is referred to as *adaptive* LSF time-domain filtering.

6. Other methods for amplitude/frequency determination

Many other schemes exist for estimating the time variation of SOAE amplitude and frequency, one of which is the Hilbert transform described in Murphy *et al.* (1995a). This method involves first performing an FFT of the entire data sample. A 30-Hz bandpass filter with 5-Hz tails centered at the SOAE frequency was applied to the complex spectrum, and then inverse transformed to produce a time-domain filtered signal that isolated the SOAE of interest. The Hilbert transformed signal and the filtered signal can be obtained simultaneously by zeroing the negative frequency components (de Boer, private communication, 1996). Upon performing a complex inverse FFT (with appropriate normalization), the real component of the time waveform will correspond to the filtered signal, and the imaginary part to its Hilbert transform. If $y_f(t)$ is the filtered signal and $y_h(t)$ is its Hilbert transform, then the amplitude and phase of $y_f(t)$ can be determined by

$$a(t)=\sqrt{y_f(t)^2+y_h(t)^2}, \tag{4}$$

$$\cos\ \varphi(t)=y_f(t)/a(t), \tag{5}$$

$$\sin\ \varphi(t)=y_h(t)/a(t), \tag{6}$$

where $a(t)$ and $\varphi(t)$ are the SOAE amplitude and phase, respectively. The instantaneous SOAE angular frequency $\omega_0(t)$ can be determined by noting that $\omega_0(t)=\dot{\varphi}(t)$, where the dot ''$\cdot$'' signifies a derivative with respect to time. This method has the same difficulty of being as prohibitively computational expensive as the quadratic peak fitting method. It also requires that a single frequency domain window be applied to the entire time-domain sample. Because of the variation of the SOAE frequency over the course of a long recording (the variability can be as much as 10 Hz), the bandpass filter width is forced to be much wider than is the case for methods in which the filter width dynamically tracks the SOAE frequency. Consequently the Hilbert transform method yields results with a significantly lower signal-to-noise ratio (typically 10 dB worse than the quadratic fit method).

Another method for the determination of frequency is the phase spectral method (Charpentier, 1986; Brown and Puckette, 1993). This method is well suited for the rapid determination of the frequency of one or more signal components of a recording, and is equal in sensitivity to the other methods considered. The most significant drawback of this method is that accurate amplitude estimates are not obtained. Because this method uses FFTs to obtain the frequency values, it is also a factor of $\log(N)/m$ slower than the LSF method described above.

7. Determining pulse rate

The blood flow measurements obtained using the protocol described above produced a highly asymmetric waveform [see Fig. 4(a)]. In this wave form, the peak positive amplitude corresponds to the maximum rate of blood flow. In the spectral domain, this asymmetric waveform is manifested by the presence of significant harmonic distortion [see Fig. 4(b)].

In the studies that follow, the pulse rate was determined using two different methodologies. For the measurements of Sec. III, a single pulse rate was assigned over the entire experiment. In this case, the pulse rate was determined from the peak value of the pulse rate fundamental using a spectral periodogram whose form is similar to that shown in Fig. 4(b). Since the higher-order harmonics also contained information regarding the fundamental frequency of the pulse rate, the peak frequencies of these harmonics were also included in the determination of the pulse rate frequency.

For Sec. IV, the pulse rate versus time was needed for each experimental run. To obtain this, a simple peak-picking

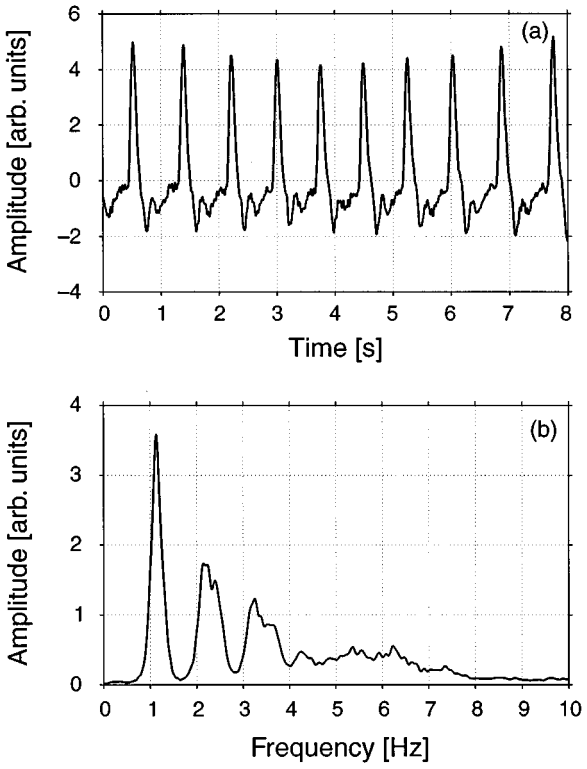


FIG. 4. (a) Time-domain pulse readout and (b) spectrum for subject JH (run 921205r). The time domain signal is highly asymmetric, leading to the significant harmonics observed in the spectral domain.

algorithm was used that selected the time of maximum blood flow for each pulse. The heartbeat frequency as a function of time was then given by one over the time difference between consecutive blood flow maxima.

II. STUDY 1

A. Methods

In a preliminary study, centered averages were performed on all the 300–600 s recordings in our data base before January 1993 with SOAEs$\gtrsim$0 dB SPL. This resulted in centered averages of 45 SOAEs in 14 ears from 10 subjects.

B. Results

Sidebands were observed in the emissions farthest from the noise floor in all tested ears (see Table I) in which the recording length was sufficiently long ($\gtrsim$300 s) to obtain a good signal-to-noise ratio.[1] Representative spectra from three of the subjects in Table I are presented in Fig. 5. The sharp peaks at the emission frequency are an artifact of the method because the largest point on the spectrum was selected as the frequency on which to center the spectrum. For signals in the presence of noise, this meant that the noise at this frequency was always additive.

The level of the noise floor and the bandwidth of the emission prevented detection of the sidebands in emissions near the noise floor. The frequency separations and amount of modulation varied between subjects. However, when sidebands were detectable, the frequency separations and the relative levels of the sidebands were the same for all emissions with detectable sidebands in that ear. This suggested that the source of modulation was common for all emissions in that ear (see Fig. 5). The level of modulation is indicated by the relative amplitude of the SOAE central frequency and the sidebands. In order to evaluate the depth of modulation (L_{mod}) the spectra have been plotted relative to the maximum level not SPL. Consequently, the level of the noise

TABLE I. Survey of results from analysis of centered averaging of a collection of recordings prior to January 1993 with spontaneous emissions greater than 0 dB SPL and recording times at least 300 s in length. Data obtained on the same day are entered in the order that they were obtained. The level of emissions from data obtained before the system was fully calibrated are only approximate (x). At times the side bands were very weak (†) or broad. In some of these instances the frequency f_{mod} and modulation depth (relative level of the sidebands L_{mod}) were poorly determined (*).

Subject	Date	Run	Ear	Length (s)	Number SOAEs	Max. level (dB SPL)	f_{mod} (Hz)	L_{mod} (dB)
BD	92/01/02	L	Left	300	4	10.8	1.02†	−15
BD	92/01/02	R	Right	300	7	11.4	0.85*†	−12
ED	92/01/02	L1	Left	300	1	9.4	1.15*	−10
JC	91/01/02		Left	374	6	11.7	1.05	−13
JH	92/03/13	R	Right	300	2	14.1	1.12	−17
JH	92/03/13	L	Left	300	2	2.7	1.03	−12
KH	91/08/23	A	Right	316	5	10.5	none	none
KH	91/08/23	B	Right	365	2	11.7	0.92*†	−15
KM	91/08/20	A	Right	300	1	2.0x	1.22	−12
LN	92/02/17	R	Right	300	3	10.5	1.20	−10
LN	92/02/17	RB	Right	300	3	10.5	1.04	−7
MOG	91/09/27	A	Left	300	1	3.5	1.0*	−8
RR	92/11/04	L	Left	300	5	10.4	1.10	−6
RR	92/11/04	R	Right	300	5	11.6	1.05	−15
RR	92/11/04	RR	Right	300	5	13.1	1.05	−16
RR	92/11/04	LL	Left	300	5	10.2	1.10	−6
SS	91/09/10	A	Left	300	3	9.5x	1.2*	−18
SS	91/09/10	C	Right	600	2	11.0x	1.2*†	−21

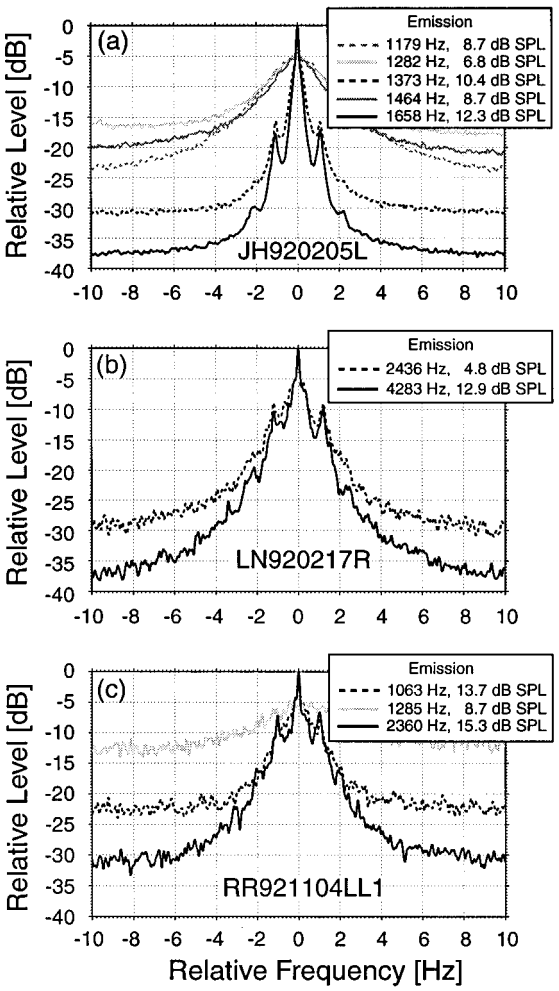


FIG. 5. Spectral line shapes of SOAEs from three human subjects: (a) subject JH; (b) subject LN; and (c) subject RR. To facilitate the comparison of the various SOAEs, the center of the emissions have been shifted to coincide with 0 Hz, and the amplitudes have been shifted so that each emission peaks at a relative power of 0 dB.

floor depends on the signal-to-noise ratio of the emissions, which, in turn, depends on the level and frequency of the emission. This was a consequence of the fact that the level of the noise floor in our system depends on frequency (see Fig. 1).

Since the spectral sidebands were of comparable spacing for all tested SOAEs with observable sidebands, a single value of f_{mod} was determined for each ear. This could be accomplished visually by setting the vertical grid spacing in the plotting program on the NeXT computer used to display the centered spectrogram to an initial estimate of the frequency spacing. By adjusting this value until the vertical grid lines coincided with the peaks of the sidebands, f_{mod} could be determined in most cases to about 0.01 Hz. In a few cases, the sidebands were either poorly resolved or very broad, and are so indicated in Table I using a dagger symbol (†). Because of this, the value of f_{mod} could not always be precisely determined, and these cases are so indicated using an asterisk (*) in Table I.

In one subject, no sidebands were observed at the beginning of a session, but appeared later in the same session. The emission frequency can change rapidly at the beginning of a session (Fig. 2; Zurek, 1981; Whitehead, 1991) and heartbeat can also be expected to reduce and stabilize as the subject relaxes in the reclining chair. The combined effects of the initial change in heartbeat and frequency make it more difficult to detect the sidebands. In most other subjects, this same pattern of increasing distinctiveness of modulation was observed on measurements obtained after both the heartbeat and the emission frequency had stabilized. In nine subjects in which more than one measurement was obtained in a single session, the frequency separation of the spectral sidebands decreased in the second recording, consistent with the expected decrease in the heartbeat rate in a resting subject.

III. STUDY 2

The data obtained in study 1 were highly suggestive of a tie between heartbeat and modulation frequency because the frequency separation of the sidebands was consistent with expected changes in heartbeat rate. The nature of the connection can only be determined by simultaneous measurement of heartbeat (blood flow) and the frequency separation of the sidebands. The relation between heartbeat rate and the frequency separation of the sidebands in the SOAE spectrum was investigated under conditions where relatively large changes in heartbeat frequency were present.

A. Methods

Recordings (300-s length) were obtained from three subjects after they rushed late into the laboratory for a study, had gone up and down stairs for a short break, when they were walking in place, and when they had been relaxed for some time in the booth. In a few instances 60-s and 180-s recordings were obtained from subject LN to evaluate the minimum length necessary for the analysis. The subjects were also asked to breathe slowly or rapidly. As before, the modulation rate of the emissions was determined by evaluating frequency separation between the center frequency of the otoacoustic emission and the peak of the sidebands. In order to optimize the detection of the sidebands, only the SOAE in each subject that had the best signal-to-noise ratio on the day of the data collection was analyzed.

B. Results

The observed modulation rate was highly correlated with pulse rate (see Fig. 6), supporting the hypothesis that sidebands in the spectrum were due to modulation by cochlear blood flow. The two points which are encircled on the figure represent data in which the frequency spacing between adjacent sidebands were poorly estimated (± 0.1 Hz) due to the short length (60 s). The unweighted linear correlation for all of the data was $r=0.982$ ($N=40$, $p<5\times10^{-10}$), implying a highly significant correlation between the measured frequency separation of the sidebands and heartbeat rate.

Sidebands in a spectrum could stem from either frequency or amplitude modulation of the emission. LSF analyses were performed to determine the frequency [see Fig. 7(a)] and amplitude [see Fig. 7(b)] fluctuations of the emissions as a function of time. The heartbeat (ac recording) is

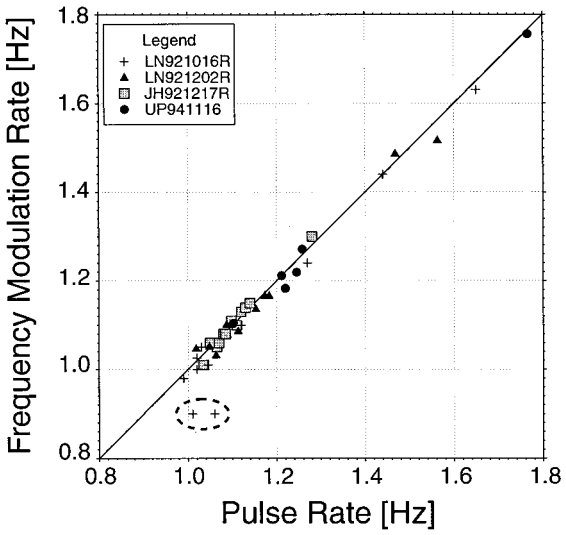


FIG. 6. Cumulative data from three pulse rate variation experiments from subjects LN (SOAE $f=4270$ Hz, $L=14$ dB SPL), JH (SOAE $f=1660$ Hz, $L=12$ dB SPL), and UP (SOAE $f=2707$ Hz, $L=-2$ dB SPL). The measurements for LN's SOAE were obtained on two separate days (10/16/94 and 12/02/94) and are displayed with different symbols (see legend) to indicate the level of reproducibility of the data. The two points which are encircled represent data in which the frequency splittings (frequency spacing between adjacent sidebands) were poorly estimated (± 0.1 Hz) due to the shorter length (60 s) of those two recordings. The linear correlation for all of the data treating all points equally was $r=0.982$ ($N=40$, $p<5\times10^{-10}$), implying a highly significant correlation between measured frequency splitting and pulse rate.

shown in Fig. 7(a) for comparison. The maximum blood flow occurred near the maximum positive value of a filtered output from the photo-plethysometer and was approximately 180° out of phase with respect to the SOAE frequency modulation. The frequency and amplitude of the otoacoustic emission were extracted from the ear canal signal with a time windows of 0.56 s using both the FFT procedure and the adaptive LSF time-domain filtering (see Sec. I). The pattern of results using the LSF (solid black line) and FFT (dashed black line) methods are essentially identical (even though the LSF is much more computationally efficient). The SOAE frequency was always a minima when the blood flow at the fingertip was maximal.

When the amplitude is extracted from 0.56-s long windows [see Fig. 7(b)] some amplitude fluctuations tied to the heartbeat remain. The 0.56-s long window gives a filter width of 1.75 Hz and the peak-to-peak frequency modulation is 1–2 Hz, consequently the frequency modulation produces an amplitude modulation of the output of the filter. The LSF procedure was thus repeated with 0.2-s windows increasing the bandwidth of the filter to 5 Hz (wider than the range of frequency fluctuations). The gray line in Fig. 7(b) obtained with the wider filter has reduced amplitude fluctuations. This optimized estimate of amplitude was used in later analyses. An eight-order polynomial was applied to the pattern of changes in frequency and level with time in order to remove the frequency drift and slow amplitude level variations over the duration of the recording. Spectral analysis of the re-

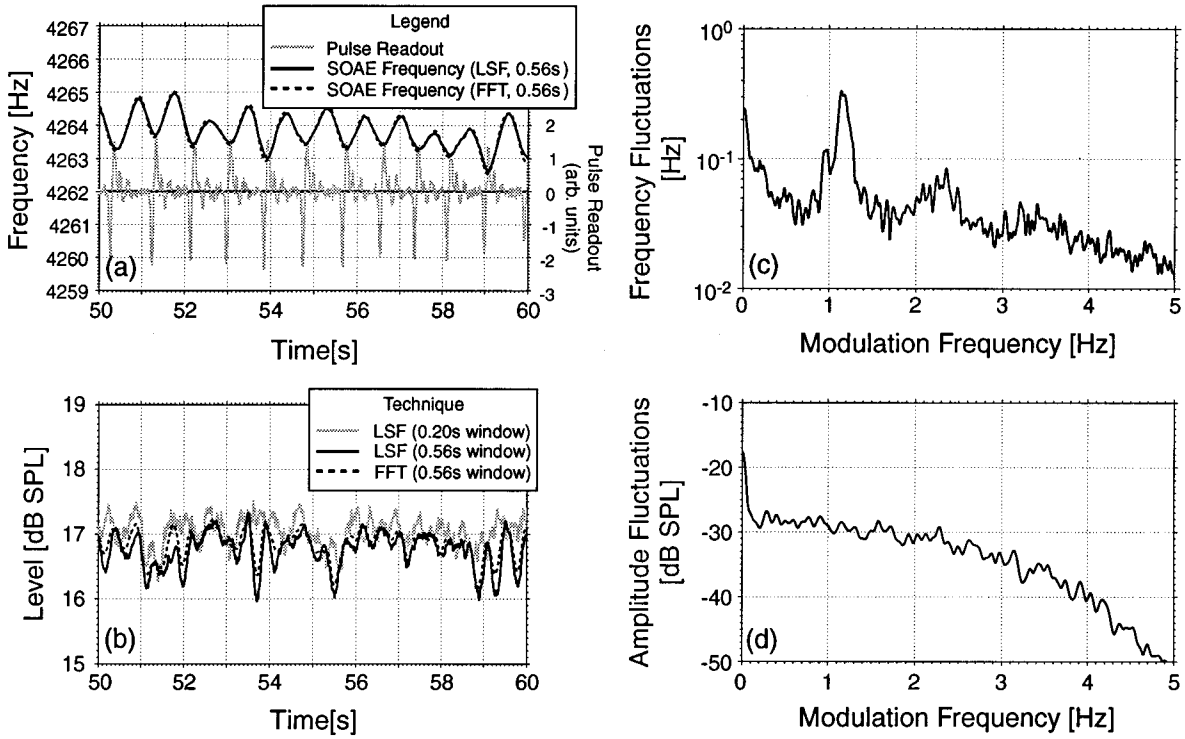


FIG. 7. (a) Instantaneous frequency and (b) amplitude of an SOAE for subject LN (run 941219r1). Spectrum of the (c) frequency and (d) amplitude fluctuations using a time window of 0.2 s. Note that the pulse readout is also displayed in (a). The large negative excursion seen in the pulse readout signal is an artifact of bandpass filtering of the raw pulse readout recording. See text for further discussion.

maining frequency fluctuations showed a distinct peak in the fluctuation spectrum at the heartbeat frequency. Higher harmonics of the heartbeat rate were also present [see Fig. 7(c)]. These reflected the asymmetrical nature of the frequency modulation shown in Fig. 7(a), where the increase in frequency was more rapid than the decrease in frequency. Once the background level of the emissions was subtracted, the spectrum of the amplitude fluctuations of the emissions was smaller than -30 dB SPL. [There is no peak detectable in Fig. 7(c).] The analysis thus indicated that the sidebands were primarily generated by frequency modulation with the frequency minima coinciding with the maximal blood flow.

IV. STUDY 3

The results from study 2 strongly linked heartbeat to the observed frequency fluctuations in SOAEs. However, it was recognized that the large changes in heartbeat rate were not needed to obtain a correlation between the two measures. In this final study, it was decided to regulate heartbeat using slow breathing, after noting that both heartbeat and the frequency modulation of the emission were modulated by the slow breathing condition in study 2. Although the peak changes in heartbeat rate were less than for study 2, the overall noise floor was less, and tracking the variation over time of heartbeat rate and of SOAE frequency and amplitude became easier.

A. Methods

Two subjects were asked to breathe very slowly (approximately 6 respirations per min) while the heartbeat rate was monitored using peak amplitude from the photoplethysometer as described in Sec. I. Data of SOAE frequency as a function of time obtained using the LSF technique (see Sec. I) were analyzed using the same procedure used to obtain estimates of the amplitude and frequency of the modulation of the SOAE frequency throughout the recording. The analyzed signals had lower noise floors than the ear canal signals from Sec. II, consequently, both amplitude and frequency of modulation could be obtained as a function of time, and directly compared to heart rate.

B. Results

Figure 8(a) and (b) shows the pattern of changes in both the heart rate and the frequency modulation rate with slow breathing for two subjects, LN and UP. LN had an emission near 4265 Hz that was well above the noise floor at 14 dB SPL, and UP had an emission near 2350 Hz with a level of 2 dB SPL. The correlation between the heartbeat rate and the rate of the frequency modulation was very good for both subjects.

UP's frequency modulation data were slightly noisier because it was harder to separate the emission from the background noise. This was a consequence of both the lower level of UP's emission, as well as the shape of the noise floor for our system. As shown in Fig. 1, the noise floor was approximately 7 dB higher at 2350 Hz than at 4265 Hz.

The data of UP are displayed in Fig. 9 as a ''scatter plot'' showing statistically independent determinations of the

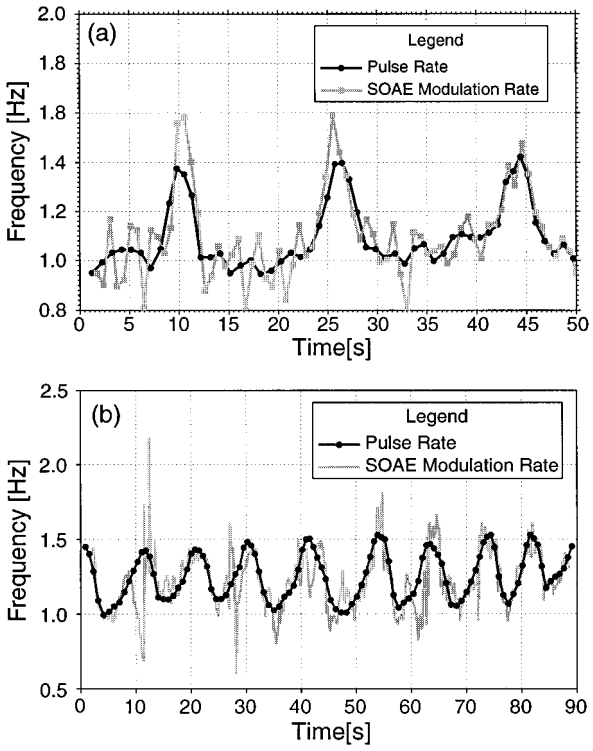


FIG. 8. Instantaneous frequency modulation rate versus pulse rate for (a) subject LN (run 941219r2), and (b) for subject UP (run 941116r3). In both cases, the heartbeat rate was modulated by ''slow breathing'' of the subject.

frequency modulation rate and heartbeat rate. The correlation of the heartbeat and the frequency modulation rate was highly significant in these data ($r=0.68$, $N=87$, $p<7\times10^{-6}$). From these results, it was concluded that it was not necessary to evaluate the emission in conditions with relatively large changes in heartbeat in order to evaluate the effect of heartbeat on otoacoustic emissions.

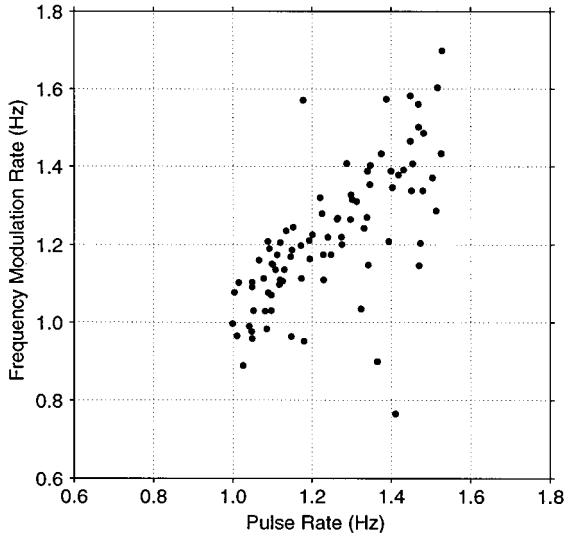


FIG. 9. Combined data from six runs for instantaneous frequency modulation rate versus pulse rate for subject UP (run 941116). The linear correlation of these data was $r=0.68$ ($N=87$, $p<7\times10^{-6}$).

V. LIMIT CYCLE MODELS OF SPONTANEOUS EMISSIONS

Various possible mechanisms by which heartbeat could give rise to amplitude and/or frequency fluctuations can be evaluated in computer simulations. The advantage of this approach over a more qualitative one is that it permits the estimation of the expected relative amplitude of the modulation of SOAE amplitude and frequency under different conditions. These results can then be used to evaluate the observation that frequency modulation is large and amplitude modulation is below the noise floor.

Many aspects of the frequency and amplitude of SOAEs have been successfully simulated by modeling the emission as a limit cycle oscillator (e.g., Talmadge *et al.*, 1991). Consequently, the effects of heartbeat modulation on the parameters of a nonlinear limit cycle oscillator were investigated. Discussion of the relationship of these results to models in which SOAEs are generated as a result of the reflection of energy from random inhomogeneities on the basilar membrane (e.g., Talmadge and Tubis, 1993) will be deferred to a future publication.

A. Limit cycle oscillator model

The forced Van der Pol–Duffing oscillator may be written in the generalized form,

$$\ddot{x}(t)+[-r+sx^2(t)]\dot{x}(t)+\omega_0^2x(t)+\omega_0^2\epsilon x^3(t)=\omega_0^2\Lambda(t), \quad (7)$$

where x refers to the pressure level of the SOAE in the ear canal, ω_0 is the natural frequency of the oscillator, r is the negative resistance that gives rise to self-oscillation, s is the nonlinear damping that stabilizes the oscillator, and ϵ is a nonlinear stiffness contribution. Finally, the forcing term $\omega_0^2\Lambda(t)$ is assumed to be a weak slowly varying signal similar to that which could be produced by modulation of cochlear pressure by blood flow. This pressure fluctuation could be a result of low-frequency acoustic stimulus or modulation of cochlear pressure through the cochlear aqueduct as suggested by Bell (1992).

Equation (7) can be approximately solved to second order using Krylov–Bogoliubov–Mitripolski (KBM) averaging (e.g., Jackson, 1989), which has the the steady-state solution

$$x(t)\cong\overline{a(t)}\cos(\overline{\omega(t)}\cdot t+\varphi_0), \quad (8)$$

$$a_0=\left(\frac{r}{4s}\right)^{1/2}, \quad (9)$$

$$\frac{\overline{a(t)}}{a_0}\cong 1-\frac{a_0^2\epsilon}{16}-\frac{2\Lambda^2(t)}{a_0^2}, \quad (10)$$

$$\frac{\overline{\omega(t)}}{\omega_0}\cong 1+\frac{3a_0^2\epsilon}{8}-\frac{r^2}{16\omega_0^2}+\frac{3}{2}\,\epsilon\Lambda^2(t). \quad (11)$$

Equation (7) contains a nonlinear stiffness component which would be consistent with the stiffness of the cochlear partition (including the stereocilia) depending on displacement. Nonlinear stiffness provides a potentially significant mechanism by which amplitude modulation could be converted into frequency modulation and vice versa. Van Dijk *et al.* (1994) showed that a broadband noise stimulus gives rise to a correlation between the noise-induced amplitude and frequency fluctuations in the oscillator amplitude. This point will be returned to in the discussion.

In order to determine to source of the effects of heartbeat, Eq. (7) was modified to include separate functions for the relevant parameters,

$$[1+a_m(t)]\ddot{x}(t)+-r[1+a_r(t)]\dot{x}(t)+s[1+a_s(t)]x^2(t)\dot{x}(t)+\omega_0^2[1+a_k(t)]x(t)+\omega_0^2\epsilon[1+a_\epsilon(t)]x^3(t)=\omega_0^2\Lambda(t). \quad (12)$$

The functions $a_m(t)$ (mass), $a_r(t)$ (negative resistance), $a_s(t)$ (nonlinear resistance), $a_k(t)$ (stiffness), and $a_\epsilon(t)$ (nonlinear stiffness) are slowly varying perturbative functions that can be systematically varied to evaluate the possible impact of the modulation of that parameter by heartbeat on the amplitude and frequency of the emission. In addition, the forcing term $\Lambda(t)$ is interpreted as a slow perturbative pressure fluctuation. With these assumptions, the KBM solution can be expanded to lowest nontrivial order in the various perturbations to find

$$x(t)\cong\overline{a^*}(t)\cos(\overline{\omega^*}(t)\cdot t+\overline{\varphi^*}), \quad (13)$$

$$\frac{\overline{a^*}(t)}{a_0}\cong 1-\frac{a_0^2\epsilon}{16}-\frac{2\Lambda^2(t)}{a_0^2}+\frac{a_r}{2}-\frac{a_s}{2}-\frac{a_0^2a_\epsilon}{16}, \quad (14)$$

$$\frac{\overline{\omega^*}(t)}{\omega_0}\cong 1-\frac{r^2}{16\omega_0^2}-\frac{r^2a_r}{8\omega_0^2}+\frac{3a_0^2\epsilon}{8}\,(1+a_r-a_s)+\frac{3a_0^2a_\epsilon}{8}+\frac{3}{2}\,\epsilon\Lambda^2(t)+\frac{1}{2}\,(a_k-a_m). \quad (15)$$

For the purpose of discussion, the quantities are introduced,

$$A_a\equiv\frac{\overline{a^*}-\overline{a}}{a_0}=-\frac{2\Lambda^2(t)}{a_0^2}+\frac{a_r}{2}-\frac{a_s}{2}-\frac{a_0^2a_\epsilon}{16}, \quad (16)$$

$$A_\omega\equiv\frac{\overline{\omega^*}-\overline{\omega}}{\omega_0}=-\frac{r^2a_r}{8\omega_0^2}+\frac{3a_0^2\epsilon}{8}\,(a_r-a_s)+\frac{3a_0^2a_\epsilon}{8}+\frac{3}{2}\,\epsilon\Lambda^2(t)+\frac{1}{2}\,(a_k-a_m), \quad (17)$$

where A_a and A_ω are the relative amplitudes of the modulation of the amplitude $a(t)$ and frequency $\omega_0(t)$, respectively, of the oscillator simulating the SOAE.

To the degree that Eqs. (13)–(15) accurately describe the influence of heartbeat on the amplitude [Eq. (14)] and frequency [Eq. (15)] of SOAEs, these equations have great predictive power in allowing one to evaluate possible physical mechanisms for the influence of heartbeat and/or blood flow on SOAEs. In particular, they allow one to explore the magnitude of the effects of modulation of one or more SOAE parameters upon SOAE amplitude and frequency.

Because of the power of Eqs. (13)–(15) in terms of evaluating the effectiveness of the various proposed mechanisms, it is important that the accuracy of these equations be numerically verified. This is because most of the results from

these equations are first-order contributions to amplitude and/or frequency modulation, and the possibility exists that higher-order corrections could significantly influence the results. A series of computer simulations were performed during the course of this study to verify different aspects of the calculations. In general, the results of these simulations confirm Eqs. (13)–(15) to a high degree of accuracy.

In studying the various mechanisms, the parameters of the oscillator were chosen to match the phenomenological values of real SOAEs, namely $\omega_0 = 2\pi \times 4000$ rad s^{-1}, $r = 100$ rad s^{-1} (e.g., Murphy *et al.*, 1995a) and $|\epsilon| \lesssim 0.01$ (e.g., van Dijk *et al.*, 1994). For numerical convenience, $a_0 = 1$ was chosen. The modulation rate was $f_{hb} = 1$ Hz ($\omega_{hb} = 2\pi$ rad s^{-1}) in all cases, and the modulation level was chosen to correspond to typical levels of modulation observed in Secs. I–III. Unless otherwise specified, the nonlinear stiffness component was set to $\epsilon = 0$.

1. Effect of an external driver

The influence of a low-frequency driving stimulation of the form $\Lambda(t) = \Lambda_0 \cos(\omega_{hb} t)$ is considered first. This low-frequency driver could arise either from acoustical stimulation in the ear canal that is transmitted via middle ear and/or bone conduction, or from pressure variations in the cerebral spinal fluid (see Sec. VI). Computer simulations were performed using Eq. (12) for values of Λ_0 ranging between 0 and 0.3. Representative simulation results are displayed in Fig. 10 for the case $\Lambda_0 \cong 0.189$.

Two immediate results are apparent from this figure. First, frequency modulation is present at a significantly reduced level (~ -100 dB) compared to amplitude modulation. This result can be understood from Eq. (15), in which no first-order contribution from $\Lambda(t)$ is present, in the absence of nonlinear stiffness. Second, the amplitude modulation occurs at $2\omega_{hb}$, which can be understood from Eq. (14), in which the external stimulus appears as $\Lambda^2(t) = \Lambda_0^2 \cos \omega_{hb}^2 t = \Lambda_0^2/2 \cdot [1 + \cos(2\omega_{hb} t)]$.

2. Modulation of linear stiffness

A number of simulations were performed in which the linear stiffness of the oscillator was modulated for values of a_k ranging from 0.001 to 0.040, and the results for $a_k = 0.0158$ are shown in Fig. 11. In this case, frequency modulation was observed at $1\omega_{hb}$ with an amplitude of 2.98 Hz (~ -63-dB modulation level). When the window length was sufficiently small (see discussion below on window length effects), the amplitude modulation was found to be less than -96 dB. These results can be understood using Eqs. (14) and (15). No first-order contribution from a_k to amplitude modulation occurs in Eq. (14), and a frequency effect that is in phase with the stiffness modulation is expected with a relative frequency modulation amplitude $A_\omega = a_k/2 \approx 0.0079$. These predictions compare favorably to the simulation result of $A_\omega = 2.98$ Hz/4000 Hz ≈ 0.0075, and were consistent with the phase of the frequency modulation observed in the simulations.

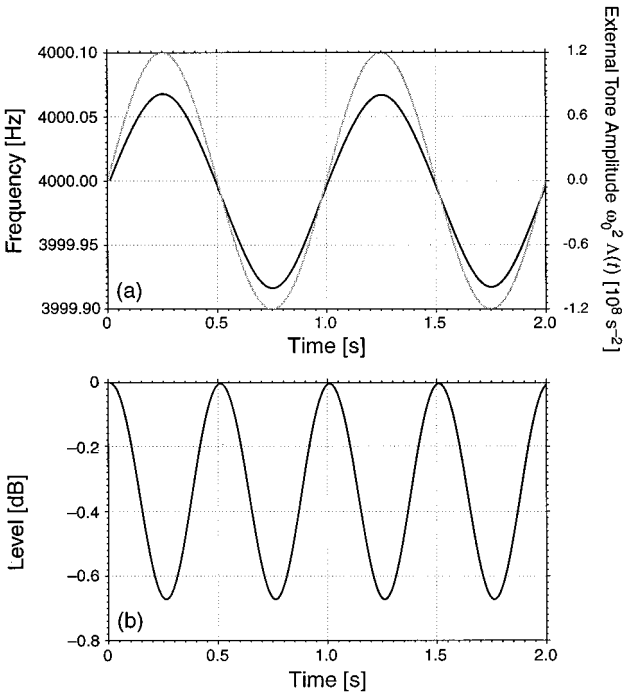


FIG. 10. (a) Instantaneous oscillator frequency (black), and (b) instantaneous oscillator amplitude of a Van der Pol oscillator for a slowly varying external tone $\omega_0^2 \Lambda(t)$ with stimulus frequency 1 Hz, and amplitude $\omega_0^2 \Lambda = 1.20 \times 10^8$ s^{-2} (corresponding to a value of $\Lambda = 0.189$). Also plotted in (a) is the instantaneous driver amplitude (gray), with the axis corresponding to this quantity being shown at right. As expected in the model calculations, the amplitude modulation occurs at twice the driver frequency. A very small frequency modulation (~ 0.7-Hz amplitude) was also noticed that was in phase with the stimulus tone. However, as expected from the absence of frequency modulation in the first-order Krylov–Bugliubov calculation, the amount of frequency modulation is significantly reduced relative to the observed amount of amplitude modulation.

3. Modulation of mass

Modulation in the mass of the oscillator was simulated for values of a_m ranging from 0.001 to 0.040, and the results for $a_m = 0.001$ are shown in Fig. 12. Frequency modulation was observed at $1\omega_{hb}$ with an amplitude of frequency modulation of 1.95 Hz (or ~ -66-dB relative modulation level), but with opposite phase to the mass modulation function. A small amount of amplitude modulation (~ -96 dB) at $1\omega_{hb}$ was also observed. These results can again be directly explained from Eqs. (14) and (15). No first-order contribution from a_k to amplitude modulation occurs in Eq. (14). A frequency modulation 180° out-of-phase with the mass modulation and a modulation of $A_\omega = a_m/2 = 0.0005$ are predicted for this case. The predicted phase was again consistent with the frequency modulation phase from the simulations, and the frequency modulation amplitude was again in good agreement with the simulation result of $A_\omega = 1.95$ Hz/4000 Hz $\approx 0.000\,49$.

4. Modulation of linear resistance

Modulation in the negative resistance of the oscillator was simulated for values of a_r ranging from 0.1 to 0.5. Even for high levels of modulation, the amplitude and frequency modulations agreed well with the calculations. In this case, Eq. (14) predicted amplitude modulation at the level of A_a

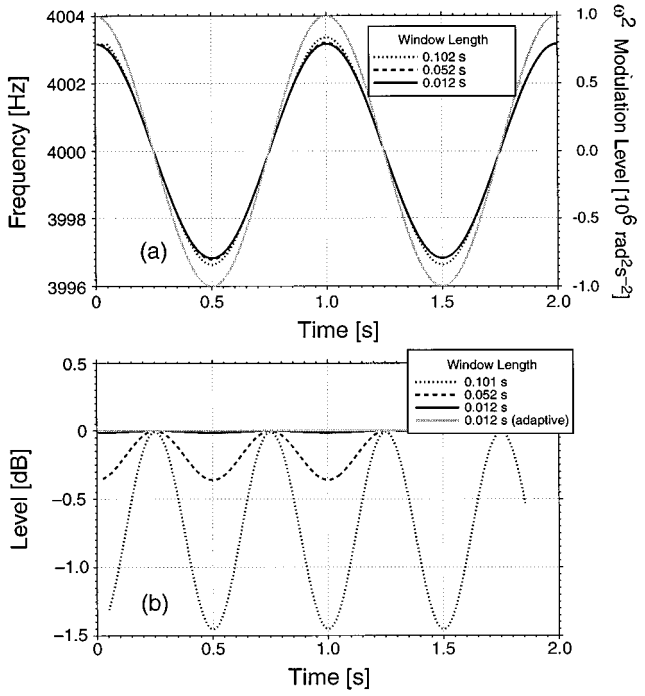


FIG. 11. (a) Instantaneous oscillator frequency (black), and (b) instantaneous oscillator amplitude of a Van der Pol oscillator for a time-varying stiffness ($f_{hb}=1$ Hz, $\omega_0^2 a_k=10^6$ rad^2 s^{-2} corresponding to $a_k\cong 0.001\,58$), as evaluated using different window sizes with the LSF technique. Also shown in (a) is the stiffness modulation level (gray), with the axis corresponding to this quantity being shown at right. The amplitude determined using the adaptive LSF method is shown in (b) for purposes of comparison. The analogous result for (a) is not shown as it is not distinguishable from the standard LSF technique when a 0.012-s window is used.

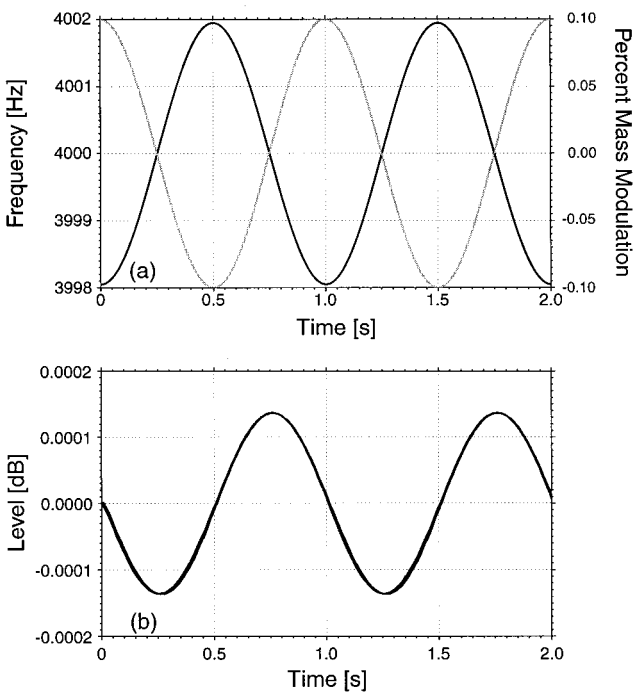


FIG. 12. (a) Instantaneous oscillator frequency (black), and (b) instantaneous oscillator amplitude of a Van der Pol oscillator with a time-varying mass ($f_{hb}=1$ Hz, $a_m=0.001$). Also shown in (a) is the percent variation in mass (gray), with the axis corresponding to this quantity being shown at right. As predicted the dominant variation is in frequency rather than amplitude. The very small fluctuations in amplitude may be artifactual, but contributions from higher-order perturbations giving rise to an amplitude modulation cannot be ruled out.

$=a_r/2$, and a second-order frequency modulation that is opposite in sign to $a_r(t)$ and reduced relative to A_a by the factor $r^2/4\omega_0^2\approx 4\times 10^{-6}$. With the choice of $a_r=0.5$, the simulation produced an amplitude modulation of $A_a=0.253$ compared with the predicted value of $A_a=0.250$. For frequency modulation, the simulation value was $A_\omega=1.02\times 10^{-6}$, and the predicted value was $A_\omega=0.99\times 10^{-6}$. The predicted and simulated phases of A_a and A_f were also in agreement.

5. Modulation of nonlinear stiffness

Simulations with a nonzero value of the Duffing nonlinearity coefficient ϵ were performed which included a number of simulations with no parametric modulation present to test the predicted effect of a nonzero $\Lambda(t)$. These were compared with simulations in which the oscillator frequency varied with changing value ϵ. The results of these simulations again compared favorably to the predictions, Eqs. (11) and (15). For example, for $\epsilon=+0.01$ and $a_\epsilon=0.005$, simulation results were obtained with $A_a=3.12\times 10^{-4}$ (180° out-of-phase) and $A_f=1.81\times 10^{-3}$ (in-phase), whereas Eqs. (14) and (15) predict $A_a=3.13\times 10^{-4}$ (180° out-of-phase) and $A_f=1.88\times 10^{-3}$ (in-phase).

6. Modulation of total damping

Finally, the case of $a_r=a_s$ is considered, which might be valid if the total damping were modulated instead of only one of either the linear or nonlinear contributions. Then, inserting this contribution into Eqs. (16) and (17), and ignoring other possible sources of modulation, gives

$$A_a=0, \tag{18}$$

$$A_f=-\frac{r^2 a_r}{8\omega_0^2}. \tag{19}$$

On face value, this model seems capable of describing the observed effects described in Secs. I–III. In practice, the value of r/ω_0 is constrained by experiment (Murphy *et al.*, 1995a), e.g., for subject LN, $r^2/\omega_0^2\approx 1.4\times 10^{-5}$. Thus even if a_r were as large as 0.5, the total frequency modulation would only be 0.06 Hz, which is more than a factor of 10 too small to account for the results reported in Secs. I–III. Furthermore, because of the manner in which Eqs. (14) and (15) were derived, the nonvanishing term in Eq. (19) is a second-order contribution, whereas Eq. (18) is zero to first order only. Consequently, a second-order term might also be present in that equation. It is found for this case from simulation results that $A_a\cong A_f$.

This model is a good example of the limitations of Eqs. (14) and (15). While they can lead to insight into the nature of the contributions to A_a and A_f from various models, this insight should be verified by simulation. It is also important in applying these equations to remember that the parameters of the model are highly constrained by data on SOAEs such as their statistical properties (e.g., Talmadge *et al.*, 1991),

their amplitude versus bandwidth relationship (e.g., Talmadge *et al.*, 1993a), the time course of their response to external stimuli (e.g., Murphy *et al.*, 1995a), and the amount of nonlinear stiffness (van Dijk *et al.*, 1994).

7. Effect of window length

In order to show the sensitivity of the analysis technique on window size, the amplitude and frequency measurements were performed for the stiffness modulation simulations for a variety of window lengths [see Fig. 11(a),(b)]. These simulations produced a limit cycle oscillator that was dominantly frequency modulated. Although a very small variation in estimated oscillator frequency was obtained using different window lengths, a significant variation in the level of oscillator amplitude was observed.

This effect is a result of the frequency-to-amplitude conversion of a frequency modulated signal when a frequency selective filter is placed near the center of the frequency modulation band, and the filter has a width that is smaller than the width of the frequency modulated signal. In this case, the signal will ''warble'' in and out of the central frequency band of the filter, resulting in a diminished amplitude output when the frequency modulated signal is outside the band of the filter, compared to when it is at the center of the frequency modulation band. However, the opposite effect—namely the transformation of an amplitude modulated signal into a frequency modulated signal resulting from the analysis window length—does not occur.

B. Embedded Van der Pol model

The phenomenological model of an SOAE given by Eq. (7) has provided an excellent description with good predictive power of the properties of isolated SOAEs (e.g., Talmadge *et al.*, 1991; Murphy *et al.*, 1995a, b). However, one must be careful to recognize that the parameters of this model do not necessarily have a direct correspondence with cochlear parameters, but rather are lumped parameters that include effects from the middle and outer ear as well, and possibly even effects from the microphone used to detect the signals.

In particular, the efficiency of the middle ear in transmitting signals from the cochlea into the outer ear is governed by the mismatch between the impedance at the base of the cochlea and that of the middle ear (driven at the stapes). Hence the level of SOAEs *as measured in the ear canal* should be expected to vary as the basilar membrane (BM) stiffness at the base of the cochlea is varied. Level effects from the variation of other BM parameters at the base are also expected, but since the impedance at the base is dominated by the BM stiffness (for the frequency range of interest in this study), effects from the variation of these parameters are expected to be highly suppressed.

It is not known whether or not the parametric modulation of a given cochlear parameter should occur uniformly throughout the cochlea. Depending on the nature of the underlying source of the variation, one might expect effects that give rise to an approximately constant relative variation or to an approximately constant absolute variation in a given cochlear parameter. This issue is approached by separately parametrizing the modulation level at the base of the cochlear parameter and the modulation level near the SOAE activity maximum. The notation a_α and $\tilde{a}_\alpha$ is used to distinguish between the relative variation of a BM parameter α at the SOAE activity maximum (a_α) and near the base of the cochlea ($\tilde{a}_\alpha$).

Following Talmadge *et al.* (1997), a simple middle ear model is employed that assumes a rigid tympanic membrane and a noncompliant incudo-stapedial joint. Under these conditions, the middle ear equation reduces to a single degree of freedom system, with the property (in the absence of a external tone in the ear canal) that

$$P_e = K_0(\omega) P_d'(0), \tag{20}$$

where P_e is the incremental pressure in the ear canal, $K_0(\omega)$ is a function of frequency that depends on the middle and outer ear but not on the cochlea [so that any variation of the parameters of the cochlea will not influence $K_0(\omega)$], $P_d(0)$ is the pressure difference between scala vestibuli and scala tympani measured at the base of the cochlea, and $P_d'(0)$ is its longitudinal spatial derivative. For frequencies $\omega \ll \omega_{\mathrm{BM}}(0)$, where $\omega_{\mathrm{BM}}(x)$ is the BM frequency at position x measured from the base,

$$P_d'(0) \cong K_1(\omega) \frac{\omega}{\sqrt{\sigma_{\mathrm{BM}}(0) S}\, \omega_{\mathrm{BM}}(0)} P_d(0), \tag{21}$$

where $\sigma_{\mathrm{BM}}(0)$ is the mass density of the BM at the base, S is the average of the cross-sectional areas of the scala vestibuli and tympani at the base, and $K_1(\omega)$ is a function of frequency that depends the middle and outer ear but not on the cochlea [so that once again variation of the parameters of the cochlea will not influence $K_1(\omega)$].

Following the above discussion, modified versions of Eqs. (14) and (16) are found,

$$\frac{\overline{a^*}(t)}{a_0} \cong 1 - \frac{\alpha_0^2 \epsilon}{16} - \frac{2\Lambda^2(t)}{a_0^2} + \frac{a_r}{2} - \frac{a_s}{2} - \frac{1}{2}\tilde{a}_k - \frac{a_0^2 a_\epsilon}{16}, \tag{22}$$

$$A_a = -\frac{2\Lambda^2(t)}{a_0^2} + \frac{a_r}{2} - \frac{a_s}{2} - \frac{1}{2}\tilde{a}_k - \frac{a_0^2 a_\epsilon}{16}, \tag{23}$$

with Eqs. (15) and (17) remaining unmodified. At this level of approximation, only a variation in BM stiffness at the base will give rise to any observable new effects. Namely, a stiffness modulation at the base of the cochlea will give rise to an amplitude modulation.

The modeling predictions of this section were tested by embedding a Van der Pol oscillator at its tonotopic location in an otherwise passive cochlear model. The basilar membrane stiffness was allowed to vary along the basilar membrane with a modulation frequency of 5 Hz, and a fixed relative BM stiffness modulation level of $a_k = \tilde{a}_k = 0.001$. Simulated SOAEs with frequencies of 1500 Hz and 4000 Hz were embedded on the basilar membrane, and a total of 5 s of simulation time was obtained. The results of this simulation were in qualitative agreement with Eq. (22).

For both SOAEs, the observed frequency modulation level was -33.1 dB relative to their respective central fre-

quencies, which is in excellent agreement with the predictions of Eq. (15). The amplitude modulation level observed was −37.6 dB (relative to its ear canal level) for the 1500-Hz SOAE and −32.7 dB for the 4000-Hz SOAE. These results agree qualitatively with Eq. (22), in that the observed relative amplitude modulation level is comparable to the relative frequency modulation level, but differ from the prediction from Eq. (22) for these simulation conditions that the relative level of amplitude modulation will be −6 dB less than the level of frequency modulation.

VI. DISCUSSION

It was established in this study that heartbeat modulates the frequency of SOAEs. However, since the SOAE recordings contain more noise in amplitude than in frequency, one cannot definitively state that there is no amplitude modulation (AM). Frequency modulation can be detected in most emissions, but the amount of modulation appears to depend on the ear being examined. This may reflect differences in the pattern of blood flow to the ear. The limit cycle models in Sec. V indicate that the pattern observed (large amounts of FM 180° out-of-phase with maximum blood flow together with minimal AM) is most easily explained in terms of a modulation of the mass of the basilar membrane.

Modulation of SOAEs has also been detected in guinea pigs (Ren *et al.*, 1995). Ren *et al.* did not look for frequency modulation in the guinea pig, but they did find amplitude modulation of SOAEs. The signal was analyzed using a 50-Hz bandwidth lock-in filter centered at the SOAE frequency. This filter would have been wide enough to prevent conversion of FM to AM in the human subjects of this study. However, SOAEs in guinea pigs are wider and less stable than in humans (Ohyama *et al.*, 1991; Brown *et al.*, 1990; Ren *et al.*, 1995) and have greater frequency modulation (Brown *et al.*, 1990). An analysis that separately evaluates both FM and AM is needed before one can establish the relative contributions of AM or FM to the sidebands seen in the guinea pig SOAEs.

The authors originally considered (Talmadge *et al.*, 1993b) three mechanisms which could lead to the modulation of emissions by the heartbeat.

(1) Acoustic energy from the pulse beat causing vibration of some structure located near the external ear or middle ear might interact with SOAEs. The heartbeat is sometimes audible in the ear canal recording and some individuals can hear broadband sounds generated by vascular malformations which can be recorded in the microphones used to measure SOAEs (Champlin *et al.*, 1990). Ren *et al.* (1995) were not able to detect the ear canal heart sounds in the cochlear microphonic signal recorded from the round window, suggesting that these acoustic stimuli play little role in the cochlea in guinea pigs. The transfer function of the middle ear would prevent much of the energy in the ear canal at these frequencies from entering the cochlea. However, there may be some unknown amount of transmission through body tissues and bone vibration. There are two unequally spaced heart sounds per heartbeat (Holtz, 1996; Ren *et al.*, 1995) and thus one would expect that modulation by the acoustic signal to be more complicated than that seen in this study. Furthermore, even if there were only one heart sound per heartbeat, the pattern of modulation is inconsistent with that expected from low-frequency acoustic stimulation. Zwicker (1983) showed that transient evoked otoacoustic emissions are reduced in amplitude for both positive and negative changes in pressure, indicating that low-frequency acoustic stimuli would produce an AM of emissions at twice the frequency of the acoustic stimuli. This is consistent with the model results in Sec. V which also indicate that that acoustic stimuli would produce AM at twice the heartbeat frequency and no FM. Since the otoacoustic emission data are modulated at a rate of once per heartbeat, the hypothesis can be rejected that acoustic stimuli are responsible for the modulation of the emissions.

(2) Cerebral spinal fluid (CSF) pressure is modulated by heartbeat (Marchbanks and Reid, 1990) and the cochlear fluids are connected to the CSF through the cochlear aqueduct. Since the cochlear perilymphatic fluid can be modulated by postural manipulation of CSF fluids (Marchbanks and Reid, 1990), and the frequency of SOAEs depends on posture, Bell (1992) hypothesized that CSF modulation by heartbeat could produce sidebands on SOAEs. Although a heartbeat modulation of the fluids in the open cochlea has been observed (Cooper and Rhode, 1992), the rigid walls of the cochlea and the characteristics of the cochlear aqueduct probably prevent the pulse from significantly modulating fluid pressure in the intact cochlea (Marchbanks and Reid, 1990; Gopen *et al.*, 1997), which is protected from rapid changes in CSF pressure. Furthermore, one would expect larger sidebands from respiration than blood flow, since respiration modulates CSF pressure at least four times more than does blood flow (Marchbanks and Reid, 1990). Sidebands consistent with an interaction of the emission with respiration were occasionally seen. These sidebands were also seen in the guinea pig (Ren *et al.*, 1995). More research is need to evaluate the potential effects of respiration on SOAEs.

When the pressure in the cochlea is increased by increasing the pressure in the ear canal or in the environment, the frequency of SOAEs is normally increased and the level reduced (reviewed in Hauser *et al.*, 1993). Such increases in air pressure operate on the inner ear through the middle ear and not by their effects on the CSF (Konradsson *et al.*, 1994). However, if the CSF pressure modifies otoacoustic emissions, the impact of the pressure changes would be expected to give rise to both frequency and amplitude modulation of the SOAEs.

(3) Blood flow through the vessel of the basilar membrane (VSBM) which runs under the tunnel of Corti could modulate the mass of the basilar membrane. The models outlined in Sec. V suggest that a change in mass is consistent with the observed data. It is estimated that the mass change needed to be in the range of 10–100 ppm, which is in agreement with the small mass variations expected in the basilar membrane due to blood flow. Although there is large interspecies variability in the extent of the VSBM (Axelsson and Ryan, 1988), the vessel is present throughout the human cochlea and thus would be expected to influence the majority of SOAEs. The vessel is not as reliably present in other species (Axelsson and Ryan, 1988). Although the VSBM is

continuous in the basal turn of the guinea pig, it is more irregular in the apical turns where the SOAEs observed by Ren *et al.* (1995) would be generated. This led them to suggest that one cannot rely on modulation by the VSBM to explain the guinea pig data. The increase in blood flow is, however, the only mechanism for increasing the mass of the basilar membrane and the VSBM is sporadically present in the apex of the cochlea even in the guinea pig. A more systematic investigation of the heartbeat modulation of a range of SOAEs in a single guinea pig ear is needed. If the modulation comes from blood flow through the VSBM, one would expect greater variability in the depth of modulation of different SOAEs in the guinea pig than is seen in human subjects.

Ren *et al.* (1995) suggested two other potential mechanisms:

(4) Electrical activity in the cochlea generated by heartbeat could influence the cochlea. The cardiac electrical events are conducted throughout the body. Electromechanical processes in the cochlea could generate an acoustical signal or modify the properties of the basilar membrane by modulating the OHC function. Although the heartbeat-related electrical signal can be recorded in the cochlea after much averaging (Ren *et al.*, 1995), the signal is tiny in comparison to other electrical events in the cochlea, reducing the probability that this is the primary determinant of the frequency modulation. If the electrical signal modulated the OHC function, it would be equivalent to modulating the negative resistance r which leads to large AM and small FM. An alternative hypothesis would be that electrical bias could modulate the linear or nonlinear stiffness of the basilar membrane due to changes in the shape of the OHC (see Sec. V and discussion below).

(5) Changes in the pressure in the cochlea as a consequence of the blood flow through the stria vascularis. Heartbeat-related pulsatile flow in the stria has recently been measured using laser Doppler flowmetry by Ren *et al.* (1995). Ren *et al.* (1995) hypothesized that this blood flow within the rigid walls of the cochlea would lead to pulsatile oscillation in cochlear pressure which might modulate SOAEs. The amount of change in volume and thus the increase in pressure in the total cochlea would be very small. Furthermore, since the stria vascularis is restricted to the scala media, this would be equivalent to introducing a low-frequency stimulus on the basilar membrane, which, as discussed in Sec. V, would be expected to produce large changes in amplitude at twice the heartbeat frequency. Increasing the pressure in both scala media and scala vestibuli would increase the stiffness of the basilar membrane and thus change both amplitude and frequency of the emission.

Finally, Sec. V suggests a sixth mechanism, although without a specific physiological underpinning:

(6) A modulation in nonlinear stiffness ϵ could give rise to large FM with unobservable AM. From Eqs. (16) and (17), it can be shown that $A_a \cong A_f/6$. For the observed amount of FM in Secs. II and III, this level of AM would not be detectable. Equally important, the values of ϵ reported in van Dijk *et al.* (1994) range from -0.01 to 0.015. For a nonlinear stiffness modulation of $\alpha_\epsilon \approx 0.5$, this would allow FM to be present at the level actually detected. Since ϵ can be either positive or negative, FM that is 180° out-of-phase will occur if $\epsilon < 0$, and FM in-phase will occur if $\epsilon > 0$.

A physiological model giving rise to modulation of nonlinear stiffness should then predict that in some subjects (but not all) that the frequency modulation would be *in-phase*, but be 180° out-of-phase with blood flow in other subjects. The current data set does not rule out such a possibility. The effective nonlinear stiffness can be determined through measurements of the cross-correlation function $C(\tau)$ between the SOAE amplitude and frequency fluctuations arising from noise perturbations of the SOAE (van Dijk *et al.*, 1994). One of the subjects (LN) was in fact a subject in the van Dijk *et al.* study. Unfortunately, C was dominated in LN's SOAE by the 1-Hz sideband (van Dijk *et al.*, 1994) and no limit was set on the amount of nonlinearity actually present in that subject's SOAE. Determination of the value of $C(\tau)$ arising *solely* from fluctuations uncontaminated by the 1-Hz sideband will require further improvements in the analysis, and will be the subject of a future study.

An awareness of the existence of these relatively large systematic frequency modulations generated by heartbeat are important for investigations of properties of SOAEs such as those of Van Dijk *et al.* (1994) on the correlation between amplitude and frequency fluctuations of OAEs. Past modeling efforts have assumed that the amplitude and frequency fluctuations are generated by wide-band noise. The effects of heartbeat modulation need to be included in such investigations.

The ability to evaluate aspects of cochlear blood flow noninvasively may have potential clinical utility (Ren *et al.*, 1995). Cochlear blood flow appears to depend on environmental factors (reviewed in Wangemann and Schacht, 1996), and otoacoustic emissions could extend the understanding of the mechanisms responsible for this dependence. Unfortunately, the analysis techniques employed in this study limit their use to individuals with relatively large SOAEs. Procedures to evaluate heartbeat modulation in individuals with smaller SOAEs are being explored.

VII. CONCLUSIONS

Careful spectral analysis revealed than most SOAEs have sidebands which are consistent with the interpretation that the SOAE is being either frequency or amplitude modulated by the heartbeat. When both measures were determined from recordings of sufficiently long duration to reduce the noise of the estimates, the correlation between the pulse rate and the distance of the sidebands from the carrier frequency was very high ($r = 0.982$).

Additional analysis of the data which permitted separated analysis of the changes in the frequency and level of the emission revealed that the frequency modulation was easy to separate from the noise floor. Even though the noise floor was significantly reduced, no amplitude modulation was detected once the analysis bandwidth was wide enough to prevent the frequency modulation from being converted into amplitude modulation.

Computer simulation of SOAEs (outlined in Sec. V) were used to examine various hypotheses as to the potential

origin of the heartbeat modulation. It is concluded that the most likely hypothesis is a change in the mass of the organ of Corti, probably through blood flow through the vessel of the basilar membrane.

Additional physiological research and computer simulations using a full cochlear model are needed to evaluate the aspects of cochlea blood flow and cochlear pressure leading to frequency modulation of emissions.

ACKNOWLEDGMENTS

We would like to thank Arnold Tubis, Pim van Dijk, Bill Brownell, Jack Cullen, Dennis McFadden, Joe Miller, Fred Nuttall, and Alex Salt, as well as the comments of two anonymous referees, for helpful discussions and comments on various stages of this manuscript. This work was supported in part by NIH-NIDCD Grant No. DC00307 and by a grant from the Showalter Trust.

APPENDIX: LEAST-SQUARES FIT FILTER METHOD FOR DETERMINING SOAE AMPLITUDE AND FREQUENCY

This filter technique is appropriate for a signal that contains either a signal with a slowly varying frequency, or one with only an approximately determined frequency. The method described below has the additional benefit that it can determine the SOAE frequency using a *linear* filter technique.

The derivation of this method starts by assuming that the signal can be written as

$$y(t)=a\ \cos[(\omega+\Delta\omega)t]+b\ \sin[(\omega+\Delta\omega)t], \tag{A1}$$

where $\Delta\omega$ is the unknown offset to the best frequency of the signal for a given time window. This equation is linearized by assuming that $\Delta\omega t\ll 1$, so that

$$y(t)\cong a\ \cos\ \omega t+b\ \sin\ \omega t-a\Delta\omega t\ \sin\ \omega t + b\Delta\omega t\ \cos\ \omega t. \tag{A2}$$

Additional simplification can be made by writing

$$c=+b\Delta\omega, \tag{A3}$$

$$d=-a\Delta\omega, \tag{A4}$$

and treating c and d as *independent* parameters. This approximation is necessary to fully linearize Eq. (A2):

$$y(t)\cong a\ \cos\ t+b\ \sin\ \omega t+ct\ \cos\ \omega t+dt\ \sin\ \omega t. \tag{A5}$$

For notational convenience, $y(t)$ is written as

$$y(t)=y(t;\boldsymbol{\alpha})=\sum_{m=1}^{4}\alpha_m F_m(t)\equiv\boldsymbol{\alpha}\cdot\mathbf{F}, \tag{A6}$$

where

$$F_1(t)\equiv\cos(\omega t),\quad F_2(t)\equiv\sin(\omega t),\quad F_3(t)\equiv t\ \cos(\omega t),\quad F_4(t)\equiv t\ \sin(\omega t), \tag{A7}$$

and

$$\alpha_1\equiv a,\quad \alpha_2\equiv b,\quad \alpha_3\equiv c,\quad \alpha_4\equiv d. \tag{A8}$$

Given a input window with K points (where K is even), and an input signal y_l, $l=0,1,2,...,L-1$ sampled at a uniform time interval Δt, the formal sum of squares is written as

$$\chi_l^2(a,b,c,d)=\sum_{k=-K/2+1}^{K/2}[y(t_k;\boldsymbol{\alpha})-y_{k+l}]^2, \tag{A9}$$

where $t_k\equiv(k+1/2)\Delta t$. In this equation, the window location within the entire recording is determined by the value of l. Inserting Eq. (A6) into (A9) gives

$$\chi_l^2(\boldsymbol{\alpha})=\sum_{k=-K/2+1}^{K/2}\left[\sum_{m=1}^{4}\alpha_m F_m(t_k)-y_{k+l}\right]^2. \tag{A10}$$

The least-squares fit technique involves finding the value of $\boldsymbol{\alpha}$ such that χ_l^2 is minimized. This can be determined by requiring,

$$\frac{\partial\chi_l^2(\boldsymbol{\alpha})}{\partial\alpha_{m'}}=2\sum_{k=-K/2+1}^{K/2}\left[\sum_{m=1}^{4}\alpha_m F_m(t_k)-y_{k+l}\right]F_{m'}(t_k)=0. \tag{A11}$$

Simplifying this equation gives the result

$$\sum_{m=1}^{4}\sum_{k=-K/2+1}^{K/2}F_m(t_k)F_{m'}(t_k)=\sum_{k=-K/2+1}^{K/2}F_{m'}(t_k)y_{k+l}. \tag{A12}$$

Equation (A12) can be recast in terms of the matrix equation,

$$\begin{bmatrix} S_{cc} & 0 & 0 & S_{cst}\\ 0 & S_{ss} & S_{cst} & 0\\ 0 & S_{cst} & S_{cctt} & 0\\ S_{cst} & 0 & 0 & S_{sstt}\end{bmatrix}\cdot\begin{bmatrix} a\\ b\\ c\\ d\end{bmatrix}=\begin{bmatrix} S_{cy_l}\\ S_{sy_l}\\ S_{cty_l}\\ S_{sty_l}\end{bmatrix}, \tag{A13}$$

where

$$S_{cc}=\sum_k\ \cos^2(\omega t_k), \tag{A14}$$

$$S_{ss}=\sum_k\ \sin^2(\omega t_k), \tag{A15}$$

$$S_{cst}=\sum_k\ t_k\ \cos(\omega t_k)\sin(\omega t_k), \tag{A16}$$

$$S_{cctt}=\sum_k\ t_k^2\ \cos^2(\omega t_k), \tag{A17}$$

$$S_{sstt}=\sum_k\ t_k^2\ \sin^2(\omega t_k), \tag{A18}$$

$$S_{cy_l}=\sum_k\ \cos(\omega t_k)y_{k+l}, \tag{A19}$$

$$S_{sy_l}=\sum_k\ \sin(\omega t_k)y_{k+l}, \tag{A20}$$

$$S_{cty_l}=\sum_k\ t_k\ \cos(\omega t_k)y_{k+l}, \tag{A21}$$

$$S_{sty_l} = \sum_k t_k \sin(\omega t_k) y_{k+l}. \tag{A22}$$

In obtaining Eq. (A13), the facts have been used that $\cos(\omega t_k)$ and $t_k \sin(\omega t_k)$ are odd functions of t_k; $\sin(\omega t_k)$ and $t_k \cos(\omega t_k)$ are even functions of t_k; and the summation over k is being performed under symmetric limits in t, to eliminate the other cross-term sums in the 4×4 matrix.

Equation (A13) has the solution

$$a = \frac{S_{cy_l} S_{sstt} - S_{cst} S_{sty_l}}{-S_{cst}{}^2 + S_{cc} S_{sstt}}, \tag{A23}$$

$$b = \frac{-S_{cst} S_{cty_l} + S_{cctt} S_{sy_l}}{-S_{cst}{}^2 + S_{cctt} S_{ss}}, \tag{A24}$$

$$c = \frac{S_{cty_l} S_{ss} - S_{cst} S_{sy_l}}{-S_{cst}{}^2 + S_{cctt} S_{ss}}, \tag{A25}$$

$$d = \frac{-S_{cst} S_{cy_l} + S_{cc} S_{sty_l}}{-S_{cst}{}^2 + S_{cc} S_{sstt}}. \tag{A26}$$

Since $\Delta\omega$ occurs linearly in both c and d, it is now an overdetermined parameter. To extract $\Delta\omega$ from c and d, the sum-of-squares is formed,

$$\chi^2(\Delta\omega) = (c - b\Delta\omega)^2 + (d + a\Delta\omega)^2, \tag{A27}$$

and minimized with respect to $\Delta\omega$, to give

$$\Delta\omega = \frac{-ca + bd}{a^2 + b^2}. \tag{A28}$$

The model given by Eq. (A6) is used in the definition of χ_l^2 in Eq. (A9) without any explicit dependence on the location of the LSF window. That is, $y(t;\boldsymbol{\alpha})$ is evaluated at t_k rather than t_{k+l}. The reason for this choice is that the evaluations of the trigonometric functions $\sin \omega t_k$ and $\cos \omega t_k$ need to be performed only once at the start of the analysis of a recording, significantly reducing the computational overhead of the LSF method. This choice for the definition of χ_l^2 requires the phase correction,

$$a' = a \cos[\omega(l+1/2)\Delta t] - b \sin[\omega(l+1/2)\Delta t], \tag{A29}$$

$$b' = a \sin[\omega(l+1/2)\Delta t] + b \cos[\omega(l+1/2)\Delta t], \tag{A30}$$

which results from the application of the usual sine and cosine addition formulae. Here a' and b' may be regarded as the coefficients which would be obtained by using the model evaluated at t_{k+l} in Eq. (A9). Finally, the SOAE amplitude A, frequency ω_0, and phase φ is,

$$A = \sqrt{a'^2 + b'^2} = \sqrt{a^2 + b^2}, \tag{A31}$$

$$\omega_0 = \omega + \frac{-ca + bd}{A^2}, \tag{A32}$$

$$\varphi = \tan^{-1}(b'/a'). \tag{A33}$$

Although the above formalism has been developed for a single SOAE, it can be readily generalized to include multiple SOAEs, and to include one or more pure tones. In this case, a matrix equation that is a generalization of Eq. (A13) is obtained. This generalized matrix builds in correlations between the assumed external tones and SOAEs, and for this reason does not suffer from the usual resolution limits of the discrete Fourier transform. Instead, the resolution and sensitivity of the LSF technique depend on factors such as the signal-to-noise ratio, the sampling rate, etc. A more complete discussion of this issue will be reserved for a future publication by one of the authors (CLT).

[1]In our original ARO report (Talmadge *et al.*, 1993b), one additional subject (BB) was included in our table. However, the recording times for this subject were only 60 s, which is too short for sidebands of the level seen in other subjects to the observable. Unfortunately this individual was not available for retesting with longer recording times.

Axelsson, A., and Ryan, A. (**1988**). ''Circulation of the inner ear: I. Comparative study of the vascular anatomy in the mammalian cochlea,'' in *Physiology of the Ear*, edited by A. F. Jahn and J. Santos-Sacchi (Raven, New York), pp. 295–315.

Bell, A. (**1992**). ''Circadian and menstrual rhythms in frequency variations of spontaneous otoacoustic emissions from human ears,'' Hearing Res. **58**, 91–100.

de Boer, E. (**1996**). Private communication.

Brown, A. M., Woodward, S., and Gaskill, S. A. (**1990**). ''Frequency variation in spontaneous sound emissions from guinea pig and human ears,'' Eur. Arch. Otorhinolaryngol. **247**, 24–28.

Brown, J. C., and Puckette, M. S. (**1993**). ''A high resolution fundamental determination based on phase changes of the Fourier transform,'' J. Acoust. Soc. Am. **94**, 662–667.

Burns, E. M., and Pitton, J. W. (**1993**). ''Time-frequency analyses of coherent frequency fluctuations among spontaneous otoacoustic emissions,'' J. Acoust. Soc. Am. **93**, 2314.

Champlin, C. A., Muller, S. P., and Mitchell, S. A. (**1990**). ''Acoustic measurements of objective tinnitus,'' J. Speech Hear. Res. **33**, 816–21.

Charpentier, F. J. (**1986**). ''Pitch detection using the short-term phase spectrum,'' in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing* (IEEE, New York), pp. 113–116.

Cooper, N. P., and Rhode, W. S. (**1992**). ''Basilar membrane mechanics in the hook region of cat and guinea-pig cochleae: Sharp tuning and nonlinearity in the absence of baseline position shifts,'' Hearing Res. **63**, 163–190.

Dallos, P. (**1996**). ''Overview: Cochlea Neurobiology,'' in *Springer Handbook of Auditory Research, Volume 8: The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 1–43.

Gopen, Q., Rosowski, J. J., and Merchant, S. N. (**1997**). ''Anatomy of the normal human cochlear aqueduct with functional implications,'' Hearing Res. **107**, 9–22.

Hauser, R., Probst, R., and Harris, F. (**1993**). ''Effects of atmospheric pressure variation on spontaneous, transiently evoked, and distortion product otoacoustic emissions in normal human ears,'' Hearing Res. **69**, 133–145.

Holtz, J. (**1996**). ''Peripheral circulation: Fundamental concepts, comparative aspects of control in specific vascular sections, and lymph flow,'' in *Comprehensive Human Physiology*, edited by R. Gregor and U. Wondhorst (Springer-Verlag, Berlin), pp. 1865–1915.

Jackson, E. A. (**1989**). *Perspectives of Nonlinear Dynamics* (Cambridge U.P., Cambridge).

Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Konradsson, K. S., Carlborg, A. H., Farmer, Jr., J. C., and Carlborg, B. I. (**1994**). ''Perilymph pressure during hypobaric conditions-cochlear aqueduct obstructed,'' Acta Oto-Laryngol. **114**, 24–29.

Long, G. R., Tubis, A., and Jones, K. L. (**1991**). ''Modelling synchronization and suppression of spontaneous otoacoustic emissions using Van der Pol oscillators: Effects of aspirin administration,'' J. Acoust. Soc. Am. **89**, 1201–1212.

Marchbanks, R. J., and Reid, A. (**1990**). ''Cochlear and cerebrospinal fluid pressure: their inter-relationship and control mechanisms,'' Br. J. Audiol. **24**, 79–187.

McFadden, D., and Pasanen, E. G. (**1994**). ''Otoacoustic emissions and quinine sulfate,'' J. Acoust. Soc. Am. **95**, 3460–3474.

Mott, J., Norton, S., Neely, S., and Warr, B. (**1989**). ''Changes in spontaneous otoacoustic emissions produced by acoustic stimulation of the contralateral ear,'' Hearing Res. **38**, 229–242.

Murphy, W. J., Talmadge, C. L., Tubis, A., and Long, G. R. (**1995a**). ''Relaxation dynamics of spontaneous otoacoustic emissions perturbed by external tones: I. Response to pulsed single-tone suppressors,'' J. Acoust. Soc. Am. **97**, 3702–3710.

Murphy, W. J., Talmadge, C. L., Tubis, A., and Long, G. R. (**1995b**). ''Relaxation dynamics of spontaneous otoacoustic emissions perturbed by external tones: II. Suppression of interacting tones,'' J. Acoust. Soc. Am. **97**, 3711–3721.

Murugasu, E., and Russell, I. (**1996**). ''The effect of efferent stimulation on basilar membrane displacement in the basal turn of the guinea pig cochlea,'' J. Neurosci. **16**, 325–332.

Ohyama, K., Wada, H., Kobayashi, T., and Takasaka, T. (**1991**). ''Spontaneous otoacoustic emissions in the guinea pig,'' Hearing Res. **56**, 111–121.

Patuzzi, R. (**1996**). ''Cochlear micromechanics and macromechanics,'' in *Springer Handbook of Auditory Research, Volume 8: The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 186–257.

Press, W. H., Teukosky, S. A., Vetterling, W. T., and Flannery, B. P. (**1992**). *Numerical Recipes in C* (Cambridge U.P., Cambridge), 2nd ed.

Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). ''A review of otoacoustic emissions,'' J. Acoust. Soc. Am. **89**, 2027–2067.

Rao, P., and Bilger, R. (**1991**). ''Spectral instabilities of spontaneous otoacoustic emissions,'' J. Acoust. Soc. Am. **90**, 2289(A).

Ren, T., Zhang, M., Nuttal, A. L., and Miller, J. (**1995**). ''Heart beat modulation of spontaneous otoaocustic emissions in guinea pig.,'' Acta Oto-Laryngol. **115**, 725–731.

Russell, I. J., and Murugasu, E. (**1996**). ''The effect of efferent stimulation and acetlycholine perfusion on basilar membrane displacement in the basal turn of the guinea pig cochlea,'' in *Diversity in Auditory Mechanics*, edited by E. Lewis, G. Long, R. Lyon, P. E. Narins, and C. Steele (World Scientific, Singapore) (in press).

Talmadge, C., Long, G. R., Murphy, W., and Tubis, A. (**1993a**). ''New off-line method for detecting spontaneous otoacoutic emissions in human subjects,'' Hearing Res. **71**, 170–182.

Talmadge, C., and Tubis, A. (**1993**). ''On modeling the connection between spontaneous and evoked otoacoustic emissions,'' in *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Hosrt, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 25–32.

Talmadge, C., Tubis, A., Long, G. R., and Piskorski, P. (**1997**). ''Modeling otoacoustic emission fine structure,'' in *Diversity in Auditory Mechanics*, edited by E. Lewis, G. Long, R. Lyon, P. Narins, and C. Steele (World Scientific, Singapore), pp. 462–471.

Talmadge, C. L., Long, G. R., and Tubis, A. (**1993b**). ''Classification and small scale structure of spontaneous emissions,'' in *Abstracts of the Sixteenth Midwinter Research Meeting of the Association for Research in Otolaryngology*, edited by D. J. Lim (Association for Research in Otolaryngology, Des Moines), p. 98.

Talmadge, C. L., Tubis, A., Wit, H. P., and Long, G. R. (**1991**). ''Are spontaneous otoacoustic emissions generated by self-sustained cochlear oscillators?'' J. Acoust. Soc. Am. **89**, 2391–2399.

van Dijk, P., Wit, H., Tubis, A., Talmadge, C. L., and Long, G. (**1994**). ''Correlation between amplitude and frequency of spontaneous otoacoustic emissions,'' J. Acoust. Soc. Am. **96**, 163–169.

Wangemann, P., and Schacht, J. (**1996**). ''Homeostatic mechanisms in the cochlea,'' in *Springer Handbook of Auditory Research, Volume 8: The Cochlea*, edited by P. Dallos, A. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 130–185.

Whitehead, M. L. (**1991**). ''Slow variations of the amplitude and frequency of spontaneous otoacoustic emissions,'' Hearing Res. **53**, 269–280.

Zurek, P. (**1981**). ''Spontaneous narrow-band acoustic signals emitted by human ears,'' J. Acoust. Soc. Am. **69**, 514–523.

Zwicker, E., and Peisl, W. (**1983**). ''Delayed evoked oto-acoustic emission and their suppression by Gaussian-shaped pressure impulses,'' Hearing Res. **11**, 359–371.

Otoreflectance of the cochlea and middle ear

Douglas H. Keefe
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 27 May 1997; accepted for publication 8 August 1997)

Otoreflectance refers to acoustic pressure reflectance measurements in the ear canal, by the use of a leak-free insertion of a probe assembly, in the frequency or time domain over a range of two or more stimulus levels. Otoreflectance includes an iso-level response indicative of the forward transfer of acoustic energy into the middle ear, and nonlinear responses indicative of the acoustic energy reflected back from the cochlea. The nonlinear otoreflectance decouples the reflected energy in an evoked otoacoustic emission (OAE) from its subsequent re-reflected energy due to the presence of the ear-canal probe. Nonlinear otoreflectance responses are extremely sensitive to probe distortion, and a double-evoked (2E) technique, previously used in evoked (OAE) measurements, is adapted for otoreflectance to solve the distortion problem. Results are obtained using a 2E stimulus set containing a set of three click stimuli delivered through a pair of sources. The corresponding sets of three pressure responses are acquired in a calibration tube, and in the ear canal, and a set of three iso-level ear-canal reflectances is calculated. An evoked OAE can be decomposed into a otoacoustic reflected pressure (ORP), and a nonlinear otoreflectance is defined by the ratio of the ORP to the initial pressure spectrum. Otoreflectance provides simultaneous measurements of middle-ear and cochlear responses, and has potential, as yet untested, for application to clinical tests for hearing impairments. © *1997 Acoustical Society of America.* [S0001-4966(97)06411-4]

PACS numbers: 43.64.Jb, 43.64.Ha, 43.64.Yp [BLM]

LIST OF SYMBOLS

CEOAE	click-evoked otoacoustic emission
DAC1	digital-to-analog converter for channel 1
DAC2	digital-to-analog converter for channel 2
OAE	otoacoustic emission
ORP	otoacoustic reflected pressure
ORP1	otoacoustic first-reflected pressure
OR	otoreflectance
$p_1(t)$	pressure waveform response to stimulus $s_1(t)$
$p_2(t)$	pressure waveform response to stimulus $s_2(t)$
$p_{12}(t)$	pressure waveform response to stimulus $s_{12}(t)$
$p_D(t)$	double-evoked pressure response
$P_i(f)$	outgoing component of pressure $P(f)$
$P_r(f)$	incoming component of pressure $P(f)$
$P^k(f)$	total pressure corresponding to a kth-order round trip reflection
$R(f)$	ear-canal pressure reflectance
$R^0(f)$	source pressure reflectance
$R^c(f)$	round-trip calibration-tube pressure reflectance
$s(t)$	the stimulus set corresponding to the presentation of $s_1(t)$, followed by $s_2(t)$, followed by $s_{12}(t)$
$s_1(t)$	stimulus 1 waveform input to DAC1
$s_2(t)$	stimulus 2 waveform input to DAC2
$s_{12}(t)$	joint stimulus waveform comprised of stimulus 1 waveform input to DAC1 and stimulus 2 waveform input to DAC2
2E	double-evoked
2CEOAE	double-click-evoked otoacoustic emission
$\Delta P_r(f)$	nonlinear ORP
$\Delta P_r^0(f)$	nonlinear ORP_1
$\Delta R(f)$	nonlinear otoreflectance

INTRODUCTION

The underlying cochlear mechanics that generate otoacoustic emissions (OAE) also produce measurable changes in such ear-canal transfer functions as acoustic impedance, acoustic admittance, and reflectance (Jurzitza and Hemmert, 1992; Allen *et al.*, 1995; Keefe, 1995; Burns and Keefe, 1997; Burns *et al.*, 1997). Such transfer functions may provide a more fundamental ear-canal description of cochlear mechanical processes, which are observable by ear-canal measurements, than measurements of sound pressure level alone. The need for a transfer function representation of OAEs is exemplified by the observation that the fine structure of evoked OAEs is influenced by the presence of acoustic reflections from the probe assembly inserted within the ear canal (Zwicker, 1990). A more fundamental representation of fine structure may be obtained by controlling for the acoustic response of the probe assembly. A transfer function measurement requires a calibration procedure that is, ordinarily, highly sensitive to measurement-system distortion and noise. The transfer function is a concept derived from linear-systems analysis, but it can be useful in nonlinear-system analysis, as long as the measurement-system distortion is controlled.

Otoreflectance (OR) refers to acoustic pressure reflectance measurements in the ear canal in the frequency or time domain over a range of two or more stimulus levels. The basic idea is to decompose both the acoustic stimulus and the response of the ear into incident and reflected signals. An evoked OAE is a reflected signal from the cochlea (Kemp, 1978), and it is therefore natural to analyze OAEs within the

framework of OR. A typical evoked OAE signal is a superposition of the reflected cochlear signal with its re-reflected signal from the surface of a probe back along the incident direction; higher-order reflections may also be present. The OR technique provides a means to separate the reflected and re-reflected sets of OAE responses.

Within the linear approximation, the ear-canal reflectance is a transfer function that characterizes the acoustic response of the middle ear (Stinson *et al.*, 1982; Hudde, 1983; Stinson, 1990; Keefe *et al.*, 1993; Voss and Allen, 1994) and that may be useful in detecting and diagnosing middle-ear impairments (Keefe and Levi, 1996). Under this approximation, the reflectance should be independent of stimulus level. This is the case for sufficiently high (and sufficiently low) stimulus levels, but, as stated above, the existence of spontaneous and evoked OAEs are evidence of level-dependent changes in reflectance. A reference to a linear transfer function of the ear can be misleading, so that such a transfer function is referred to herein as an *iso-level* transfer function of the ear. This makes explicit the dependence of the transfer function on stimulus level.

An iso-level transfer function, such as pressure reflectance, then becomes a measure of middle-ear response, but it is implicit that the transfer function depends on the iso-level response of the cochlea (e.g., its input impedance at the oval window) as well as the contribution from OAEs. Similarly, an evoked OAE measurement is not only influenced by the cochlear response, but also by the middle-ear pathway in both the forward and reverse directions as well as the acoustic reflection properties of the probe assembly. This probe response is controlled in the acoustic transfer function measurement by means of a calibration procedure.

A single-tube calibration measurement technique for acoustic reflectance measurements was developed (Keefe, 1996b), and applied to the measurement of otoreflectance responses using a probe assembly with a single acoustic source driven by a digital-to-analog converter (DAC) and a single microphone coupled to an analog-to-digital converter (Keefe, 1995). One conclusion from this study was that measurement-system distortion was a significant limitation in low-level, nonlinear OR measurements.

This problem was addressed in two stages. The first stage was based on the recognition that nonlinear OR measurements have many similarities to click-evoked otoacoustic emission measurements (CEOAE), and, more generally, transient-evoked otoacoustic emissions (TEOAE). Thus the problem of measurement-system distortion in OR could be approached by focusing on the problem of distortion in CEOAEs, the drawbacks of which typically result in the use of time gating to eliminate the initial portion of the CEOAE that is contaminated by probe distortion. This problem was solved for CEOAEs by introducing a double-evoked (2E) stimulus/response technique based on the use of two DACs, each coupled to its own acoustic source (Keefe, 1996a, 1997b; Keefe and Ling, 1997). The second stage is to adapt the 2E technique to OR measurements. This report describes the theory of OR measurement and presents selected results.[1]

I. MEASUREMENT THEORY

The theory underlying the otoreflectance measurement technique is described in this section in three parts. The first part describes the technique used to measure an iso-level acoustic reflectance. This differs from the technique used in other investigations of ear-canal reflectance (Keefe *et al.*, 1993; Voss and Allen, 1994) insofar as only a single-tube response is needed to calibrate the reflectance system. The second part summarizes pertinent aspects of the 2E CEOAE technique. The third part describes the otoreflectance technique using 2E stimulus sets.

A. Reflectance calibration

While a mathematical time-domain description of the theory underlying the single-tube reflectance method is available (Keefe, 1996b),[2] this section presents the theory in a manner relevant to ear-canal measurements. When mathematical relations are introduced, time-domain (frequency-domain) variables are represented by lower-case (upper-case) symbols. Any such pair of variables are related by the Fourier transform (FT).

Before an ear-canal reflectance can be measured, it is necessary to determine the characteristics of the initial signal, which is delivered by the acoustic source(s) within the probe assembly, and the reflectance of the probe assembly itself. These can be inferred from the pressure signal measured by inserting the probe assembly into one end of a cylindrical calibration tube of length L that is closed at its far end. If the tube is sufficiently long, the short-duration incident signal generated by the probe source is separable from its first-reflected signal, which is the transformation of the incident signal by propagation down the tube, reflection from its closed end, and propagation back to the probe microphone.

The distance x along the tube axis is measured such that $x=0$ lies at the probe surface, which is the location of the short-duration acoustic flow injected into the tube, and $x=L$ is the closed end. A small microphone, which is located at $x=0$ within the probe assembly, measures a total initial pressure waveform $p^0(t)$ in response to this input flow, which is incident in the direction away from the probe toward the closed end. This wave is followed in terms of its subsequent reflections.

It travels a distance L down the tube, is reflected at the closed end, and travels a distance L up the tube arriving back at the probe surface as the signal $\hat{p}_r^1(t)$. The superscript 1 and subscript r denote that this is the first-reflected wave. The nominal round-trip time is $2L/c$, such that c is the free-space phase velocity of sound. By causality, $\hat{p}_r^1(t)$ is zero for times earlier than $2L/c$. The corresponding spectra are related by the round-trip calibration-tube pressure reflectance $R^c(f)$ defined by

$$R^c(f)=\frac{\hat{P}_r^1(f)}{P^0(f)}. \tag{1}$$

This tube reflectance is entirely determined as a function of frequency using an accurate model of viscothermal propagation in cylindrical tubes, which depends only on its length,

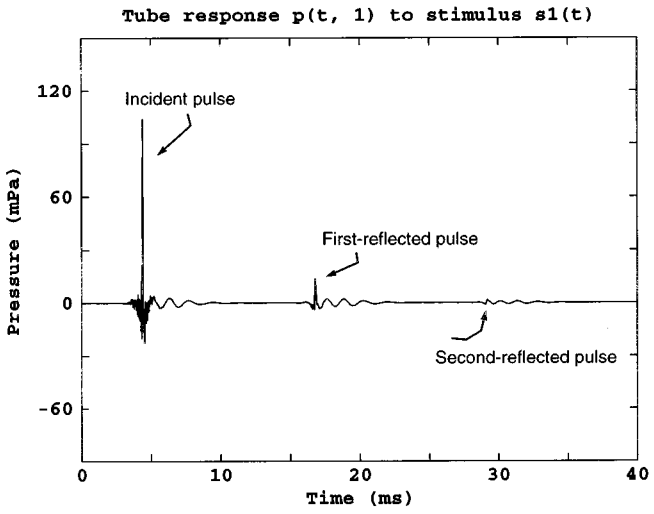


FIG. 1. The calibration-tube pressure waveform is illustrated in response to an acoustic click stimulus representing $s_1(t)$. The incident, first-reflected, and second-reflected pulses lie completely within the measurement duration. Higher-order reflected pulses from the opposite closed end of the tube are allowed to decay before another acoustic click is generated to prevent temporal aliasing of the response.

cross-sectional area S, and frequency.[3] It is expressed by

$$R^c(f)=e^{-\Gamma(2L)}, \tag{2}$$

where the complex propagation wave number Γ is a function of L, S, and f that depends on viscous and thermal conduction parameters of air (Keefe, 1984). The length and area are known for any particular calibration tube, such that the area is chosen to be similar to an ear-canal area. In the absence of viscothermal losses, the reflectance is a pure time delay $e^{-j2\pi f(2L/c)}$ corresponding to the round-trip delay, expressed using the unit imaginary number j. Thus any error in the tube length is approximately equal to a phase error in the reflectance, a property utilized below.

The return wave $\hat{p}_r^1(t)$ is reflected at $x=0$ from the probe surface and generates a newly incident wave $\hat{p}_i^1(t)$ that proceeds onward toward the closed end again. The total first-reflected pressure measured by the microphone is $\hat{p}^1(t)=\hat{p}_r^1(t)+\hat{p}_i^1(t)$. The source reflectance $R^0(f)$ determines the relationship between the incoming and outgoing pressure waves at the probe surface by

$$R^0(f)=\frac{\hat{P}_i^1(f)}{\hat{P}_r^1(f)}. \tag{3}$$

The first-reflected pressure spectrum is

$$\hat{P}^1(f)=\hat{P}_r^1(f)+\hat{P}_i^1(f)=[1+R^0(f)]\hat{P}_r^1(f). \tag{4}$$

It is assumed that the duration of the input flow is sufficiently short that the initial microphone signal $p^0(t)$ has completely decayed before the onset of the first-reflected microphone signal $\hat{p}^1(t)$. This separability criterion ensures that the microphone signal can be decomposed into the initial and first-reflected signals, and allows calculation of their respective FTs, $P^0(f)$ and $\hat{P}^1(f)$. An example in Fig. 1 of separability in a measured tube response shows $p^0(t)$ as the ''Incident pulse'' and $\hat{p}^1(t)$ as the ''First-reflected pulse.'' It follows from Eqs. (1) and (4) that

$$\hat{P}^1(f)=[1+R^0(f)]R^c(f)P^0(f), \tag{5}$$

so that the source reflectance is calculated in terms of known quantities by

$$R^0(f)=\frac{\hat{P}^1(f)-R^c(f)P^0(f)}{R^c(f)P^0(f)}. \tag{6}$$

A typical ear-canal probe assembly is comprised of a foam or plastic structure within which are capillary tubes conducting the acoustic signals. There may be some uncertainty in the length of the tube after insertion of the probe assembly into the calibration tube. Moreover, any temperature change lends uncertainty by influencing the phase velocity of sound, and thus the round-trip delay time. Such a perturbation can be represented in terms of a temperature-dependent change in the equivalent length of the calibration tube. In either case, a small change in tube length predominantly modifies the phase of the tube reflectance $R^c(f)$ and thus produces a phase error in the measurement of the source reflectance $R^0(f)$ via Eq. (6).

In practice, the source reflectance phase averaged across frequency should be close to zero—any linear trend in phase with frequency would, to good approximation, be equivalent to a modification of the tube length. Such a linear trend can be removed by adjusting the equivalent length of the calibration tube. The tube length is set to an initial value indicative of the actual length, and subsequent calibrations adjust this length until the average phase of the source reflectance is in the neighborhood of zero. Small variations in the source reflectance phase within this neighborhood produce no significant changes in the measured ear-canal reflectance. This single-tube technique has the advantages of simplicity and ease of usage without sacrificing accuracy, as has been demonstrated by comparing the results from this technique to the source reflectance calculated by an alternative, multi-tube calibration technique (Keefe, 1997a). The parsing of the overall round-trip phase into an adjustment of the equivalent tube length is intimately related to causality requirements. The tube length in the model might be adjusted based on formal causality requirements on the source reflectance, noting that the tube reflectance trivially satisfies causality via the separatibility criterion. This approach was not used because the measured reflectance has a limited bandwidth whereas accurate use of causality in the frequency domain requires a full-bandwidth measurement of reflectance.

The primary outputs from the calibration include the initial pressure spectrum $P^0(f)$ and source reflectance $R^0(f)$, while secondary outputs include the equivalent tube length L and the round-trip tube reflectance model $R^c(f)$ calculated using this length.

1. Reflectance measurement in the ear canal

After calibration is completed, the probe assembly is inserted into the ear canal of the subject and a pressure response, with waveform $p(t)$ and corresponding spectrum $P(f)$, is measured using the same electrical stimulus as in

the calibration-tube response measurements. The ear-canal pressure reflectance to be measured is $R(f)$.

The ear-canal pressure is the superposition of the individual reflections from the ear canal, which is terminated by the middle ear and cochlea, and from the probe surface. The ear-canal pressure reflectance is a total, round-trip reflectance, which incorporates all reflections from sites distributed along the ear canal, middle ear and cochlea. Any outgoing component from the probe is reflected in accordance with the ear-canal reflectance, and an incoming component to the probe is reflected by the source reflectance. The pressure measured by the microphone can be formally decomposed into an outgoing pressure $P_i(f)$ and incoming pressure $P_r(f)$, such that

$$P(f)=P_i(f)+P_r(f). \tag{7}$$

This method of analysis is in the time domain, although each individual reflection is expressed as a spectrum in the frequency domain to avoid use of convolution.

Just as for the calibration tube, the initial acoustic flow produces an initial pressure $P^0(f)$. This outgoing wave is converted into an incoming wave via the ear-canal reflectance so that the initial approximations, denoted by superscripts 0, of the outgoing and incoming spectra are

$$P_i^0=P^0, \quad P_r^0=RP^0, \tag{8}$$

The explicit dependence of each variable upon frequency is suppressed. This pair of spectra are not separable according to the above separability criterion because their respective waveforms overlap in time.

The incoming spectrum component is re-reflected at the probe surface, and converted to a new component of the outgoing signal with spectrum $R^0P_r^0=R^0RP^0$, and this outgoing component is converted to an incoming signal by the ear-canal reflectance with spectrum $RR^0P_r^0=RR^0RP^0$. The first-order pair of spectra is

$$P_i^1=P^0\{1+R^0R\}, \quad P_r^1=RP^0\{1+R^0R\}. \tag{9}$$

By induction, it is clear that each successive component is obtained by the product R^0R of the reflectances. The incoming and outgoing spectra after m pairs of reflections are

$$\begin{aligned} P_i^m&=P^0\{1+R^0R+[R^0R]^2+\cdots+[R^0R]^m]\},\\ P_r^m&=RP^0\{1+R^0R+[R^0R]^2+\cdots+[R^0R]^m]\}. \end{aligned} \tag{10}$$

All the multiple reflections are taken into account in the limit of large m, and each of the corresponding sums converge in this limit to a finite result as long as $|R(f)R^0(f)|<1$. The total outgoing and incoming pressure spectra obtained in this limit are

$$P_i=\frac{P^0}{1-R^0R}, \quad P_r=\frac{RP^0}{1-R^0R}, \tag{11}$$

so that the total microphone pressure spectrum is

$$P=P^0\left(\frac{1+R}{1-R^0R}\right). \tag{12}$$

The only unknown in this equation is the ear-canal reflectance, which is calculated at an arbitrary frequency by

$$R(f)=\frac{P(f)-P^0(f)}{R^0(f)P(f)+P^0(f)}. \tag{13}$$

This is the desired result for measuring an ear-canal reflectance. This relation was previously derived without the use of the inductive summation procedure (Keefe, 1996b), but the present derivation lends insight into the temporal sequence of acoustic reflections within the ear canal.

The technique is iso-level because the incident pressure amplitude is maintained at a constant level. If the level of this incident, or ''input,'' pressure amplitude is modified, then it is straightforward to calibrate the system at each new level and measure an ear-canal pressure response at each corresponding new level. The result is a method to measure otoreflectance. One still needs to control for measurement-system distortion, which is provided by the 2E technique summarized below.

B. Double-evoked stimulus/response sets and TEOAEs

The basic goal in the 2E technique is to control for measurement-system distortion and thereby obtain a more accurate representation of the cochlear response evoked by an acoustic stimulus presented in the ear canal. This is best accomplished by delivering the stimulus set using two DACs, each coupled to a separate acoustic source, as is customary in distortion-product OAE measurements. The 2E stimulus set is comprised of a set of three elementary stimuli $\{s_1(t),s_2(t),s_{12}(t)\}$ of equal duration. The 2E pressure in the ear canal is measured in response to this stimulus set, and valid responses are ensemble averaged through multiple stimulus presentations. This pressure response is partitioned into three elementary responses $\{p_1(t),p_2(t),p_{12}(t)\}$, corrsponding to each of the elementary stimuli $\{s_1(t),s_2(t),s_{12}(t)\}$, respectively.

The linear response of the ear or coupler is eliminated by forming the double-evoked (2E) pressure response $p_D(t)$ defined by

$$p_D(t)=p_{12}(t)-[p_1(t)+p_2(t)]. \tag{14}$$

This 2E response is a measure of nonlinear distortion in the set of responses to the 2E stimuli.

In the double-source variant of the 2E technique used herein, $s_1(t)$ is output from the first DAC1 and coupled to a first acoustic source, and $s_2(t)$ is output from the second DAC2 and coupled to a second acoustic source. The third elementary stimulus $s_{12}(t)$ is created from the simultaneous presentation of $s_1(t)$ output from DAC1 and $s_2(t)$ output from DAC2. It follows that the electrical stimuli satisfy

$$0=s_{12}(t)-[s_1(t)+s_2(t)]. \tag{15}$$

In the otoreflectance measurements to be described, the $s_2(t)$ waveform is chosen to be identical to the $s_1(t)$ waveform except for a possible change in amplitude, and is constructed such that the corresponding pressure responses $p_2(t)$ and $p_1(t)$ in the calibration-tube approximate acoustic impulses in the incident portions of their response. This is slightly different than using a click stimulus for $s_2(t)$ and $s_1(t)$, in that the impulse response of each the acoustic sources is

controlled in the stimulus design. If the acoustic sources are identical, and if $s_2(t)$ is equal in amplitude to $s_1(t)$, then $s_{12}(t)$ is twice the amplitude of $s_1(t)$ or $s_2(t)$, i.e., 6 dB higher in level. The general property, of which this choice of stimuli is an example, is that the change in stimulus amplitude is created using a pair of sources that are each excited at a constant amplitude, so that the measurement-system distortion particular to each DAC and source is held constant, and subsequently eliminated, using the 2E subtraction. The distortion pressure $p_D(t)$ in the ear canal is the double-click-evoked otoacoustic emission (2CEOAE).

Previous results have demonstrated that: (1) the dominant contributor to measurement-system distortion in CEOAE measurement systems is that due to nonlinearities in the single acoustic source outputting the click; (2) the use of two acoustic sources reduces measurement-system distortion by approximately 30 dB; and (3) the use of the 2E technique eliminates the need for time gating of the CEOAE response, which potentially extends the high-frequency limit of click-evoked OAEs (Keefe and Ling, 1997). These properties are shared in the otoreflectance technique.

C. Otoreflectance

This section describes how the 2E stimulus set is applied to otoreflectance measurements. A slightly modified notation is adopted such that the first argument denotes the time or frequency variable, and a second argument denotes the kth elementary stimulus, where k takes on the values 1, 2, or 12, corresponding to the corresponding electrical stimulus s_1, s_2, or s_{12}, respectively. A subscript denotes the incident or reflected component.

For each of the three electrical stimuli, data are collected in the single calibration tube, and the incident pressure and the source reflectance of the ER10C probe are measured using the acoustic reflectance technique described earlier. The stimuli are interleaved in a manner similar to that used in a previous study (Keefe and Ling, 1997), so that a substantively simultaneous calibration is achieved for each of the three stimulus conditions by three independent applications of the acoustic reflectance technique described above. After calibration, data are collected in the ear canal, and the reflectances of the ear are measured for each of the three stimulus conditions, again substantively simultaneously.

The probe assembly is placed in the ear, the electrical stimulus $S_1(f)$ is input to acoustic source 1, and the ear-canal pressure $P(f,1)$ is measured. This measured pressure can be decomposed into a total incident pressure $P_i(f,1)$ and reflected pressure $P_r(f,1)$ by

$$P(f,1)=P_i(f,1)+P_r(f,1). \tag{16}$$

This is equivalent to Eq. (7), but the dependence on stimulus level is explicit. A similar traveling-wave decomposition is defined for the responses to stimuli $S_2(f)$ and $S_{12}(f)$. The (pressure) reflectances—$R_1(f)$ at stimulus level 1, $R_2(f)$ at stimulus level 2, and $R_{12}(f)$ at stimulus level 12—are defined by

$$\begin{aligned} R_1(f)&=\frac{P_r(f,1)}{P_i(f,1)}, \quad R_2(f)=\frac{P_r(f,2)}{P_i(f,2)},\\ R_{12}(f)&=\frac{P_r(f,12)}{P_i(f,12)}. \end{aligned} \tag{17}$$

The corresponding reflected pressure responses are thus

$$\begin{aligned} P_r(f,1)&=R_1(f)P_i(f,1),\\ P_r(f,2)&=R_2(f)P_i(f,2),\\ P_r(f,12)&=R_{12}(f)P_i(f,12). \end{aligned} \tag{18}$$

With reference to each stimulus level k, the reflectance $R_k(f)$ is calculated using Eq. (13), and the total incident pressure $P_i(f,k)$ and reflected pressure $P_r(f,k)$ are calculated using Eq. (11). In this sense, three separate measurements are made of the reflectance at each of three levels, each with its own calibration.

The double-evoked otoacoustic reflected pressure response (ORP) is defined by

$$\Delta P_r(f)=P_r(f,12)-[P_r(f,1)+P_r(f,2)]. \tag{19}$$

Whereas an OAE pressure includes both incident and reflected pressures, the ORP is a function of the reflected pressures and has the property that it differs from zero only if the system tested is nonlinear. This may be advantageous for systems such as the ear in which the primary nonlinearity arises from the reflected component of the response, whereas probe distortion is primarily associated with the generation of the incident component. Any probe (and other measurement system) distortion produced independently by acoustic source 1 or acoustic source 2 (and DAC1 or DAC2) is cancelled in the ORP, so that the deterministic nonlinear response is due only to the reflected response of the ear. Although the term ''emission'' is avoided because the representation is in terms of a reflected cochlear signal rather than an emitted cochlear signal, it is the case that the ORP is the total-reflected component of the correspondingly evoked OAE.

The interpretation of the ORP depends on the specific choice of s_1 and s_2. Suppose $s_2(t)$ is a copy of $s_1(t)$ rescaled in amplitude by ϵ:

$$s_2(t)=\epsilon s_1(t), \tag{20}$$

In the experiments to be discussed, $s_1(t)$ is chosen so as to generate an acoustic impulse, but it may alternatively be selected to be a chirp or other waveform.

A *nonlinear otoreflectance* $\Delta R(f)$ is defined as the ratio of the ORP pressure response to the incident pressure at the reference stimulus $S_{12}(f)$:

$$\Delta R(f)=\frac{\Delta P_r(f)}{P_i(f,12)}. \tag{21}$$

It is calculated from

$$\begin{aligned} \Delta R(f)=&\left[\frac{1}{1+\epsilon}R_{12}(f)-\frac{P_i(f,1)}{P_i(f,12)}R_1(f)\right]\\ &+\left[\frac{\epsilon}{1+\epsilon}R_{12}(f)-\frac{P_i(f,2)}{P_i(f,12)}R_2(f)\right]. \end{aligned} \tag{22}$$

The nonlinear otoreflectance is a complex quantity whose magnitude exhibits fine structure characteristic of evoked OAEs. The corresponding *iso-level otoreflectance* is chosen to be the iso-level pressure reflectance measured at the highest stimulus level $R_{12}(f)$, for which the saturating nonlinearity associated with OAEs has the smallest effects.

Two ranges of values for the gain ϵ of source 2 relative to source 1 are of particular interest: (i) $\epsilon=1$, and, (ii) $\epsilon \ll 1$. For case (i), the acoustic outputs from sources with similar sensitivities, taken without loss of generality to be equal sensitivities after suitable calibration, are equal, except for differences in distortion produced. When $s_{12}(t)$ is applied, the acoustic signal in the ear canal is approximately doubled in amplitude, corresponding to a change in SPL of 6 dB. It follows that R_1 and R_2 are independent measurements of the same iso-level reflectance $R_<$ (the subscript $<$ denotes the reflectance measured at the lower level, and subscript $>$ denotes the reflectance at the s_{12} level), which might be calculated as the mean value:

$$R_<(f) = \frac{P_i(f,1)}{P_i(f,12)} R_1(f) + \frac{P_i(f,2)}{P_i(f,12)} R_2(f) \approx [R_1(f)+R_2(f)]/2. \quad (23)$$

The nonlinear otoreflectance in Eq. (22) would simplify to

$$\Delta R(f) = R_>(f) - R_<(f), \quad (24)$$

which is a simple difference in reflectance measured at two different levels. In practice, the general relation of Eq. (22) is used. In case (ii), with $\epsilon \ll 1$, it may be possible to measure the differential, i.e., nonlinear, otoreflectance with low distortion, but a reduction in ϵ for constant s_1 amplitude also reduces the signal-to-noise ratio of s_2.

The ORP is based on the decomposition of a TEOAE into reflected and incident components, and represents the total reflected component of the TEOAE. This is a step closer to the underlying model of the TEOAE as representing reflected energy from the cochlea. The pressure signal measured by the probe microphone in response to a click stimulus involves multiple numbers of reflections from the cochlea and middle ear, as well as from the probe assembly. The ORP includes the first-reflected component from the eardrum, but also the second-reflected component that propagates to the probe surface, reflects therefrom, and propagates back to the eardrum and cochlea, where it is reflected back to the probe microphone. Higher-reflected components include multiple round trips between the probe surface and the cochlea. This multiplicity of round-trip paths in the ORP signal may be significant, particularly when the nonlinear component is isolated, insofar as the second- and higher-reflected components are subject to much less saturation due to their drastically reduced incident levels. By analogy with Eq. (8), it is possible to define a first-reflected otoacoustic reflected pressure (ORP_1) that may be regarded as the first-reflected component of the total ORP.

The total reflected pressure $P_r(f,k)$ for the kth elementary stimulus, introduced in Eq. (16), is expressed by analogy with Eq. (11) in terms of the initial ear-canal pressure response $P^0(f,k)$ to the kth stimulus by

$$P_r(f,k) = \frac{R_k(f)P^0(f,k)}{1-R^0(f)R_k(f)}. \quad (25)$$

The iso-level ORP_1 is defined as the initial-reflected component of the pressure response at the kth stimulus level as follows:

$$P_r^0(f,k) = R_k(f)P^0(f,k). \quad (26)$$

It depends linearly on both the initial pressure spectrum and the kth iso-level otoreflectance. The initial pressure spectra $P^0(f,k)$ are measured directly from the calibration-tube measurements. The nonlinear ORP_1 is defined using the 2E subtraction procedure as

$$\Delta P_r^0(f) = P_r^0(f,12) - [P_r^0(f,1) + P_r^0(f,2)]. \quad (27)$$

The ORP_1 is expected to differ from the ORP due to the action of a multiple reflection filter $M_k(f)$, defined for the kth elementary stimulus condition by

$$M_k(f) = \frac{1}{1-R^0(f,k)R_k(f)}. \quad (28)$$

This filter is so named because it quantifies the influence of the multiple reflections in the ear canal, i.e., the total and initial reflected pressures are related by

$$P_r(f,k) = M_k(f)P_r^0(f,k). \quad (29)$$

An alternative definition of nonlinear otoreflectance, denoted as $\Delta R^0(f)$, may be constructed as the ratio of the ORP_1 with the initial incident pressure response to the s_{12} stimulus as follows:

$$\Delta R^0(f) = \frac{\Delta P_r^0(f)}{P^0(f,12)}. \quad (30)$$

For the case with $\epsilon=1$ and ideal equivalent acoustic sources in the absence of measurement-system distortion, this nonlinear otoreflectance reduces to the same limiting value as that of Eq. (24).

II. METHODS AND CALIBRATION RESULTS

Data were acquired at a sample rate of 32 kHz using a stimulus set $s(t)$ comprised of $s_1(t)$, $s_2(t)$ and $s_{12}(t)$. While the duration of each of these three stimuli was 42.6 ms, the nonzero portion of each stimulus was only a few milliseconds. An Etymotic ER-10C probe was used for all measurements. The click stimulus was designed in preliminary analyses to produce a reasonably flat pressure spectrum from 500 to 10 000 Hz for a short-duration, incident pulse that was measured in the calibration tube. Each DAC output a voltage signal at a level controlled by a Tucker–Davis attenuator at a setting of either 0 or -10 dB, which was then delivered to each of the acoustic sources of the probe assembly. The stimulus sets were outputted repetitively by the DACs. The calibration tube used in these experiments had a diameter of 0.802 cm, area $S=0.505$ cm^2, and length $L=212.8$ cm. The tube was closed at its far end. For calibration-tube measurements, additional buffers of silence were interposed between $s_1(t)$, $s_2(t)$, and $s_{12}(t)$ of sufficient duration that the pressure response in the tube decayed to the noise floor before

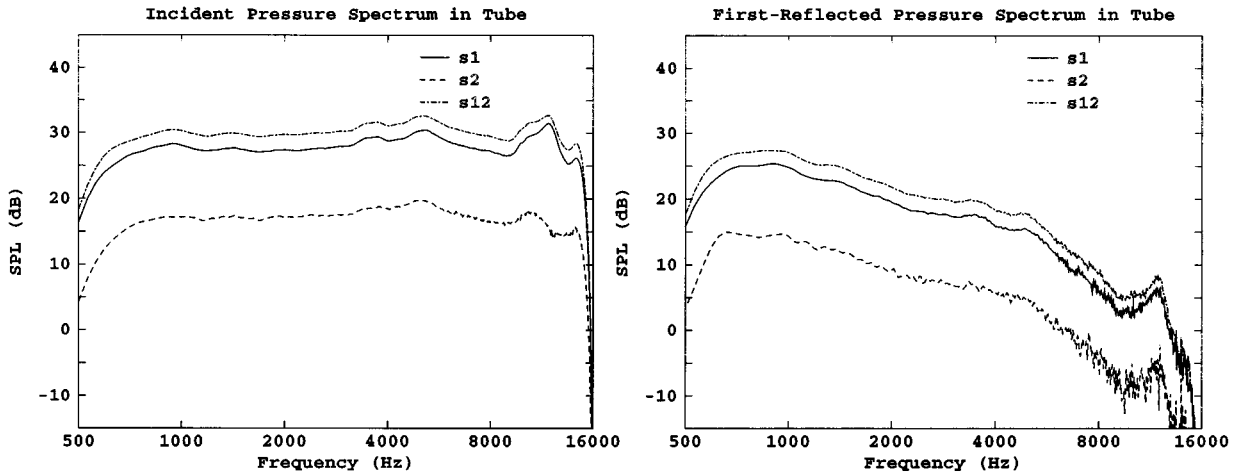


FIG. 2. The pressure spectra of each of the incident components (left panel), and first-reflected components (right panel) of the responses in the calibration tube are shown corresponding to each of the three elementary stimuli in the 2E stimulus set $\{s_1, s_2, s_{12}\}$. The responses to s_1 are the spectra of the incident and first-reflected pulses, respectively, shown in Fig. 1.

the onset of the succeeding stimulus set. No such buffers were required, or used, in the ear-canal measurements.

The microphone signal after pre-amplification was high-pass filtered using an analog filter (Krohn–Hite model 3343) whose cut-on frequency was 500 Hz, and digitally sampled during data acquisition. The validity of each acquired data buffer was judged based on a real-time intermittent noise-rejection technique, which calculates the dissimilarity of pairs of data buffers (Keefe and Ling, 1997). The analog filter was needed to attenuate the amplitude of intermittent noise at frequencies below the measurement range of interest, which would otherwise have dominated the response of the noise-rejection technique. Although this was only a problem for the ear-canal measurements, it was essential to include the filter during acquisition of the calibration-tube response in order to achieve a proper reflectance calibration. The threshold for rejecting dissimilar responses as invalid was varied for each test run so that, at the final setting, ear-canal data were acquired with only a small number of rejections. The ear-canal response was a time average of 32 valid data buffers, with an overall acquisition time as short as 4 s.

The pressure waveform measured in the calibration tube in response to the stimulus s_1 is illustrated in Fig. 1. The time origin is arbitrary. The round-trip time for a pulse to propagate down the tube and back is $2L/c = 12.4$ ms. The incident pulse is thus clearly separable from the first-reflected pulse and the second-reflected pulse (delayed by $4L/c$). The reflected pulses are not simply attenuated, but also changed in shape, due to the frequency-dependent viscothermal damping of the sound wave at the cylindrical walls. Only the incident and first-reflected pulses are required to calibrate otoreflectance. The calibration tube needs to be sufficiently long that this separability is achieved, but short enough so that the amplitude of the first-reflected pulse is sufficiently above the noise floor.

The low-frequency ringing in each pulse is due to the impulse response of the analog high-pass filter, an effect that was not taken into account in the preliminary stimulus design. The actual pressure waveform did approximate an acoustic impulse in the ear canal, and it is only the filtered electrical output from the microphone pre-amplifier that possessed the ringing tail in its response. It would be advantageous to measure the filter impulse response independently and de-convolve it from responses such as the ones obtained herein.

Separability in time of the incident and first-reflected pulses was used to calculate the source reflectance of the ER10C probe, and the incident and reflected pressure spectra (Keefe, 1996b). These spectra are illustrated in Fig. 2 in response to the stimulus set $\{s_1, s_2, s_{12}\}$ for which s_2 was attenuated by 10 dB relative to s_1, i.e., $\epsilon \approx 0.31$.

As expected from its design, the stimulus set has a reasonably constant incident pressure spectrum in the tube from 500 to 10 000 Hz (left plot, Fig. 2). However, the first-reflected pressure spectra decrease in SPL rapidly with increasing frequency above 1000 Hz, and there are problems with measurement noise above approximately 6000 Hz, particularly for the low-level s_2 signal. This noise introduces error into the reflectance calibration, so that the remaining analyses are plotted only up to 6000 Hz. Because of this noise problem, subsequent results were analyzed only for the case where the level of s_2 was the same as that of s_1, i.e., for $\epsilon = 1$.

The ER10C probe forms a complex acoustic termination to an incoming wave. It has a dense foam structure through which three capillary tubes couple the two sources and the microphone to the ear canal into which it is inserted. The pressure reflectance magnitude and phase of the ER10C probe are plotted in Fig. 3. There are three independent measurements of the source reflectance on these plots, corresponding to each of three stimuli in the 2E stimulus set. The agreement is excellent from 1000 to 5300 Hz. The probe has a resonance just above 2 kHz, and its pressure reflectance magnitude is approximately 0.6, which corresponds to an energy reflectance of 0.36 (energy reflectance is the square of the pressure reflectance magnitude). This means that 64% of the energy of an incoming sound is absorbed by the probe at this frequency. The probe reflectance is mainly resistive and

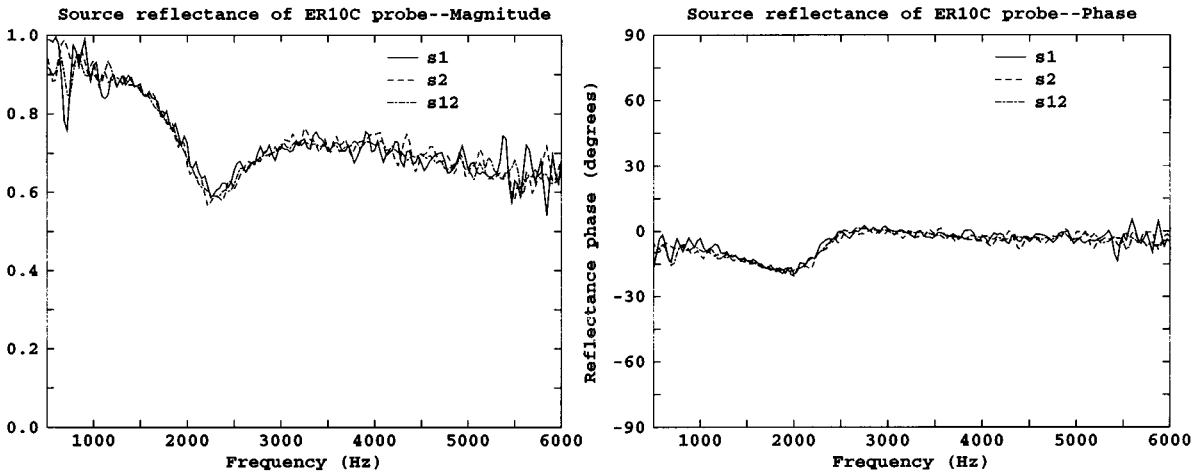


FIG. 3. The magnitude (left panel) and phase (right panel) of the source pressure reflectance of the probe assembly are illustrated, measured for each of the three elementary stimuli in the 2E stimulus set.

absorbs on the order of 50% of the incoming sound energy above 2 kHz. The acoustic response of the probe is a significant factor in reflectance and OAE tests, and changing probe design can be expected to change its acoustic reflecting properties, and thus the OAE levels measured.

III. OTOREFLECTANCE RESULTS

Responses were measured in two adult male subjects with normal hearing. The ear tested for subject 1 was known to generate spontaneous otoacoustic emissions (SOAE) and the ear tested for subject 2 had no SOAEs.

Using the stimulus s_1 at 0 dB, the waveform measured in ear canal of subject 1 is shown in Fig. 4. This response may be directly compared to the calibration-tube waveform in Fig. 1 measured using the same stimulus. The peak pressure in the ear canal is approximately 140 mPa and attenuates rapidly to small levels.

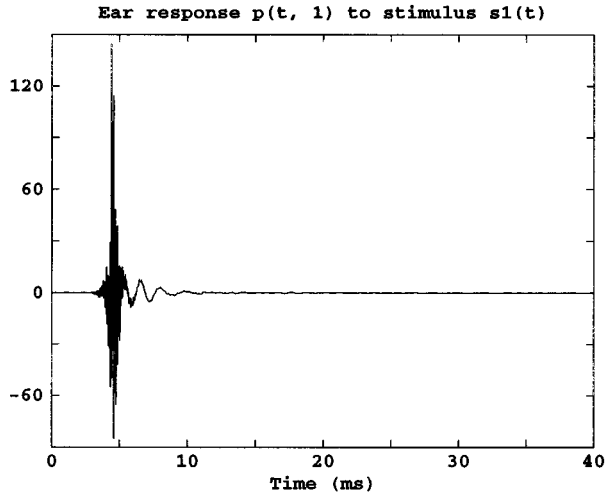


FIG. 4. The ear-canal pressure waveform (mPa) is illustrated in response to an acoustic click stimulus representing $s_1(t)$. This waveform may be compared to the corresponding calibration-tube waveform in Fig. 1 measured using the same electrical stimulus.

The double-evoked otoacoustic emission (2CEOAE) waveform for this subject, calculated using Eq. (14) had a clear onset pattern in its response.[4] The corresponding 2CEOAE spectrum was calculated as the discrete Fourier transform (DFT) of this waveform and is plotted in the left panel of Fig. 5 using a solid line for the stimulus set in which the attenuator levels of s_1 and s_2 are each 0 dB. The fine structure is replicable, with strong peaks near 2, 3.4, and 4.5 kHz. An estimate of noise is plotted using a dashed line and represents the distortion spectrum calculated using Eq. (14) based on the pressure responses measured in the calibration tube. The 2CEOAE SPL is larger than the distortion SPL in the tube at all frequencies in the 1000–6000 Hz range, except for isolated notches in the fine structure of the 2CEOAE. The right panel of Fig. 5 illustrates the otoacoustic reflected pressure spectrum (ORP) calculated using Eq. (19). For the ear-canal responses, the ORP level exceeds the 2EOAE level in the 5–6 kHz range, and the 2-kHz peak in the 2CEOAE is reduced in the ORP. A large reduction in the tube reflected pressure spectrum is evident.

The pair of plots in Fig. 6 show the iso-level pressure reflectance magnitudes (solid lines) in response to stimulus s_1 for subject 1 and subject 2. The minimum iso-level reflectance magnitude for subject 2 is approximately 0.6 at 2.7 kHz, corresponding to an energy reflectance of 0.36. The amplitude of the fine structure associated with the iso-level response is much smaller for subject 2 than for subject 1.

The nonlinear otoreflectance calculated using Eq. (22) is illustrated for each subject in the dashed lines. The fine structure of the nonlinear otoreflectance is similar to the ORP fine structure, as expected from Eq. (21). The nonlinear otoreflectance tends to be larger in amplitude in subject 1 than in subject 2, and the frequency spacings between relative maxima and minima are similar to those of the click-evoked OAEs in the same subjects.

The ORP_1 responses for subjects 1 and 2 are illustrated in Fig. 7. The top plot shows the initial pressure spectrum, the middle plot shows the iso-level ORP_1 due to the action of the middle ear as terminated by the cochlear load, and the

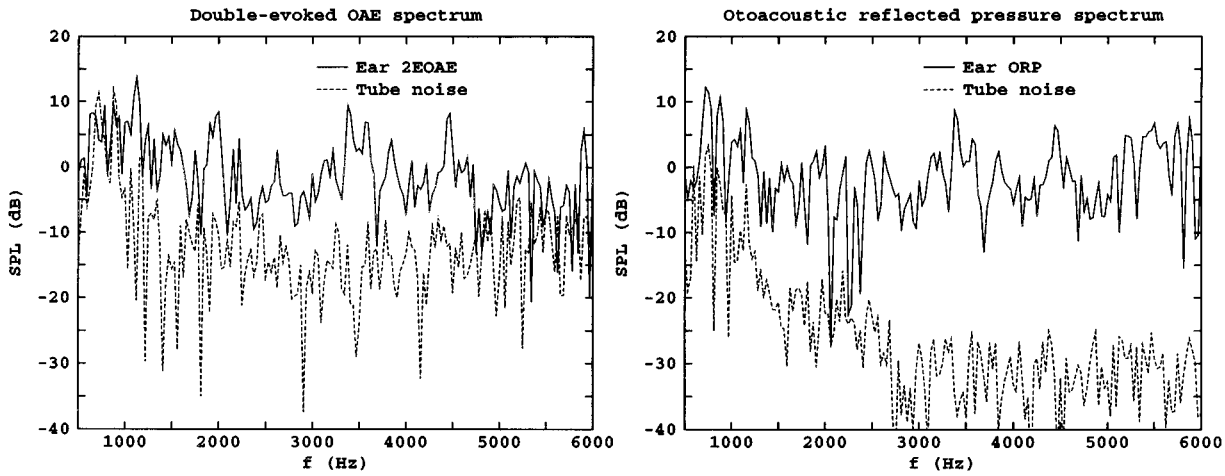


FIG. 5. The solid lines shows the double-evoked CEOAE spectrum (left panel) and the ORP spectrum (right panel) for subject 1. The dashed lines in the left and right panels show the corresponding pair of distortion spectra measured in the calibration tube.

bottom plot shows the nonlinear ORP_1 with its cochlear-induced fine structure.

The multiple reflection filter function $M_1(f)$ is plotted in Fig. 8 for subject 1, and shows a plateau level of -2 dB in the vicinity 2.0–5.4 kHz and increasing levels up to 5 dB at lower frequencies, and nearly 2 dB at frequencies up to 6 kHz. The phase is negative below approximately 3.5 kHz, and positive above. Although not plotted, the corresponding function $M_{12}(f)$ was very similar except for reduced fine structure levels, the origin of which is the fine structure associated with the ear-canal reflectance. The filter function for subject 2 has a similar shape.

IV. DISCUSSION

Otoreflectance provides a unified description of the reflecting properties of the middle ear and cochlea in response to both iso-level and nonlinear stimuli. Such a simultaneous measurement is useful in investigations aimed at cochlear functioning, because the forward transfer of power into the middle ear may be calculated in terms of the iso-level measurement (Keefe *et al.*, 1993), and the influence of the reflecting properties of the probe assembly on the magnitude of the cochlear reflections is quantified.

There were no cases observed for which the energy reflectance $|R(f)|^2$ exceeded unity in the neighborhood of SOAEs, as was observed by Burns *et al.* (1997) for subjects with large SOAEs. This is not surprising because the present study used much higher spectral levels and no particular subject criterion with respect to SOAEs. Their study varied stimulus amplitudes over a range of 25 dB whereas the stimulus amplitude in the present study was varied only by 6 dB.

Regarding the total reflected pressure $P_r(f,12)$ as an iso-level ORP, the multiple reflection filter can be viewed as the ratio of the iso-level ORP and the iso-level ORP_1. These results quantify the influence of the standing wave in the ear canal on the SPL associated with OAEs. The ORP_1 eliminates the multiple internal reflections contained in the ORP. The corresponding difference between the nonlinear ORP

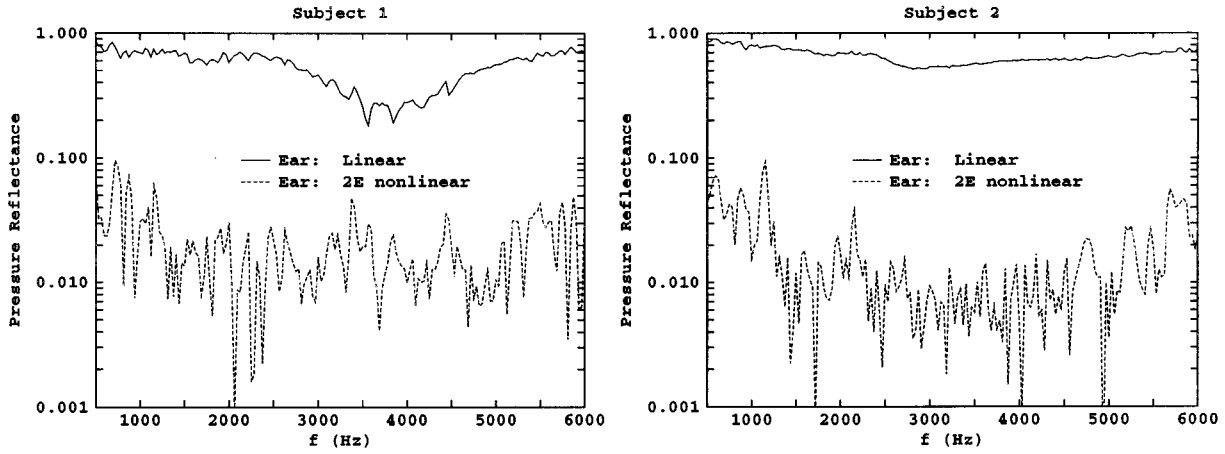


FIG. 6. OR responses are illustrated for subjects 1 and 2 in the left and right panels, respectively. The solid lines show the iso-level reflectance magnitude using the highest-amplitude click stimulus s_{12}. The dashed lines show the nonlinear OR.

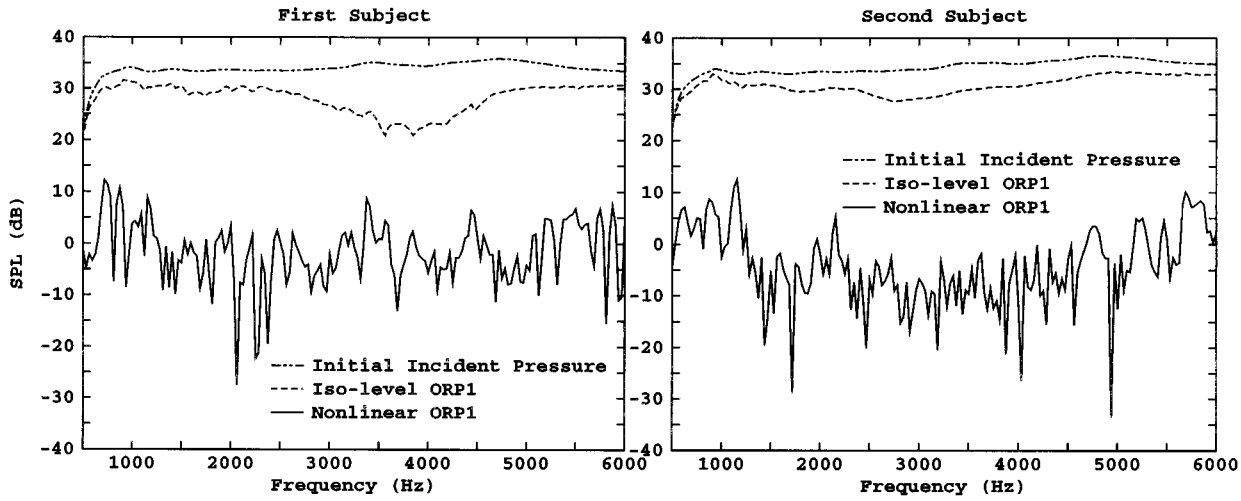


FIG. 7. ORP_1 responses for subjects 1 and 2 in the left and right panels, respectively. The top (long-short-short dashed) lines shows the SPL for the initial spectrum $P^0(f)$. The middle (dashed) lines shows the SPL for the iso-level ORP_1 $P_r^0(f,12)$ for the response to $S_{12}(f)$. The bottom (solid) lines shows the SPL for the nonlinear ORP_1 $\Delta P_r^0(f)$.

and ORP_1 was calculated using Eqs. (19) and (27). This level difference was negligible at all frequencies between 500 and 6000 Hz for both subjects.

This result is relevant to the nonlinear nature of the cochlear reflectance. In a sequence of publications culminating in Zweig and Shera (1995), a cochlear-based model of reflectance has been developed that is formally similar to the ear-canal-based model used in this report. While reviewing earlier research, Zweig and Shera observed that as the stimulus level is decreased from high to moderate levels, nonlinear contributions to the cochlear reflectance become more important, and that as stimulus levels are further decreased to the neighborhood of threshold levels, the cochlear reflectances become linear. Their discussion is based on a sinusoidal excitation model, but it has implications for any type of stimulus, including a click stimulus. They explicitly construct the multiple internal reflections in the low-level limit that the cochlear reflectance is linear, and arrive at a power series representation of reflectance of which Eq. (10), or its summed version, Eq. (12), are equivalents.

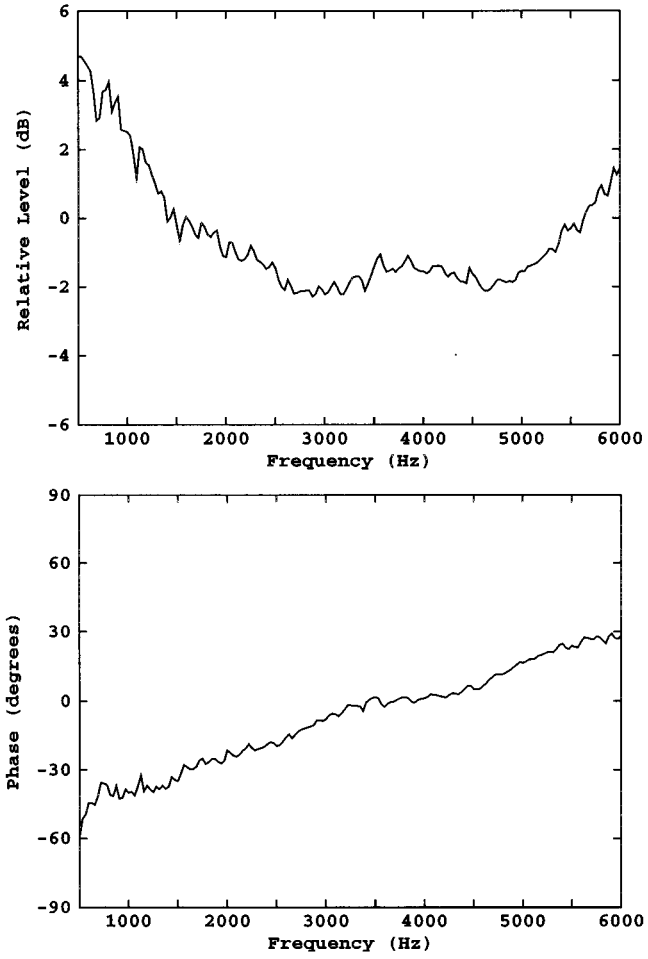


FIG. 8. The term $M_1(f)=1/(1-R^0R)$ is plotted for subject 1 based on the responses to $S_1(f)$: (a) relative level in dB; (b) phase in degrees.

Their assumption of level independence is imposed so that each instance in Eq. (10) of R or powers of R is the same reflectance function. Each higher-order reflection has a progressively lower amplitude weighting proportional to $|R^0R|^m$ for the mth reflection. If the reflectance did vary with stimulus level, then each instance of R^m corresponding to m internal reflections, would correspond to a different set of amplitude-dependent functions.

When a click stimulus is used, the initial wave has the largest amplitude and its cochlear reflectance is in the suppression regime for which the reflectance is nonlinear. The first-reflected pressure arriving back at the probe microphone is proportional to this reflectance, as expressed in Eq. (26). After subsequent re-reflection from the probe surface, the pressure wave is redirected back into the cochlea, but at a reduced amplitude, as expressed by the rightmost term in the incident pressure in the top of Eq. (9). Since its level is reduced, the corresponding cochlear reflectance converges to the linear regime with increasing numbers of internal reflections. The observed equality of the nonlinear ORP and ORP_1 in these experiments is consistent with the theory that the multiple internal reflections associated with the click stimu-

lus are functioning in the low-level linear regime of the cochlear response. However, this question remains unresolved. Investigation is needed whether there is any artifact associated with the OR measurement technique. In particular, the iso-level ear-canal reflectance is evaluated using Eq. (13) based on the representation in Eq. (12), which assumes that the reflectance is the same for each order of the internally reflected wave.

V. CONCLUSIONS

Otoreflectance (OR) provides information on sound transmittance into the middle ear via the iso-level pressure reflectance magnitude, because the transmitted energy from the incident wave is proportion to $1-|R(f)|^2$. Nonlinear OR provides information on cochlear dynamics via the otoacoustic reflected pressure and the otoreflectance pressure. These responses are obtained using in a single ear-canal measurement with an overall duration of approximately 4s, but longer averaging times may be needed to increase the signal-to-noise ratio at lower and higher frequencies. Nonlinear OR at various stimulus levels provides data on the level dependence of cochlear reflections, which can be related to predictions from cochlear models. The ORP and ORP_1 are new types of evoked OAE responses with well-defined relationships to ear-canal reflectance. These OR results are based on the use of pairs of acoustic clicks with no relative time delay. Other stimulus sets exist with interesting properties (e.g., pairs of chirps, time delay). The use of 2E stimulus/response sets in OR measurements controls for measurement-system distortion. The OR calibration requires only a single calibration tube, resulting in a simple calibration procedure for the user. A technique based solely on acoustic ear-canal measurements to simultaneously screen and diagnose middle-ear and cochlear impairments would have value in clinical applications. Future tests of OR in this regard are needed on extensive groups of subjects with normal hearing and well-defined hearing impairments.

ACKNOWLEDGMENTS

Robert Ling was of great assistance in developing the software used in this study. Michael Gorga and Stephen Neely made helpful comments on the manuscript. This research was partially supported by NIH Grant No. P01 DC00520.

[1]Some results described in this report have been presented elsewhere in preliminary form (Keefe, 1997c).

[2]Appendix A of Keefe (1996b) described an alternative reflectance measurement technique to the standard technique described in the main body of the article, and stated that its use in the frequency-domain measurement of reflectance might be preferable to the standard technique. It is not. The difference is that the standard technique uses only a relatively small time interval in the calibration-tube waveform to calculate the source reflectance, whereas the alternative technique uses the entire calibration-tube waveform. It was assumed that the use of the entire waveform might lead to improved accuracy, but this has been tested and the accuracy is lessened. The reason is that the alternative technique simply adds more noise to the system.

[3]There is a small thermal correction to the ideal closed-end reflection assumed in the theory described. It can be neglected because the calibration tube is sufficiently long that the wall losses along the tube dominate over the thermal loss at the closed end.

[4]This click-evoked OAE waveform response is plotted in Fig. 15 of Keefe and Ling (1997) and discussed therein.

Allen, J. B., Shaw, G., and Kimberley, B. P. (**1995**). ''Characterization of the nonlinear ear canal impedance at low sound levels,'' in Abstracts of the 18th Midwinter Research Meeting of the Association for Research in Otolaryngology, p. 190.

Burns, E. M., and Keefe, D. H. (**1997**). ''SOAEs and power transfer in the middle and external ears of children and adults,'' in Abstracts of the 20th Midwinter Research Meeting of the Association for Research in Otolaryngology, p. 168.

Burns, E. M., Keefe, D. H., and Ling, R. (**1997**). ''Energy reflectance in the ear canal can exceed unity near SOAE frequencies,'' J. Acoust. Soc. Am. (in press).

Hudde, H. (**1983**). ''Measurement of the eardrum impedance of human ears,'' J. Acoust. Soc. Am. **73**, 242–247.

Jurzitza, D., and Hemmert, W. (**1992**). ''Quantitative measurements of simultaneous evoked otoacoustic emissions,'' Acustica **77**, 93–99.

Keefe, D. H. (**1984**). ''Acoustical wave propagation in cylindrical ducts: Transmission line parameter approximations for isothermal and nonisothermal boundary conditions,'' J. Acoust. Soc. Am. **75**, 58–62.

Keefe, D. H. (**1995**). ''Otoreflectance system for linear and nonlinear response measurements of the middle and inner ear,'' in Abstracts of the 18th Midwinter Research Meeting of the Association for Research in Otolaryngology, p. 5.

Keefe, D. H. (**1996a**). ''New evoked cochlear responses: Double chirp-evokedy distortion products and double click-evoked otoacoustic emissions,'' J. Acoust. Soc. Am. Suppl. 1 **99**, 2562(A).

Keefe, D. H. (**1996b**). ''Wind-instrument reflection function measurements in the time domain,'' J. Acoust. Soc. Am. **99**, 2370–2381; Erratum: J. Acoust. Soc. Am. **100**, 3985 (1996).

Keefe, D. H. (**1997a**). ''Comparison of acoustic reflectance measurement techniques,'' J. Acoust. Soc. Am. **101**, 3114(A).

Keefe, D. H. (**1997b**). ''Double-evoked otoacoustic emissions. I. Measurement theory and nonlinear coherence,'' J. Acoust. Soc. Am. (submitted).

Keefe, D. H. (**1997c**). ''Nonlinear otoreflectance responses using double-evoked stimulus sets,'' in Abstracts of the 20th Midwinter Research Meeting of the Association for Research in Otolaryngology, p. 166.

Keefe, D. H., Bulen, J. C., Arehart, K. H., and Burns, E. M. (**1993**). ''Ear-canal impedance and reflection coefficient in human infants and adults,'' J. Acoust. Soc. Am. **94**, 2617–2638.

Keefe, D. H., and Levi, E. (**1996**). ''Maturation of the middle and external ears: Acoustic powerbased responses and reflectance tympanometry,'' Ear and Hearing **17**, 361–373.

Keefe, D. H., and Ling, R. (**1997**). ''Double-evoked otoacoustic emissions. II. Intermittent noise rejection, calibration and ear-canal measurements,'' J. Acoust. Soc. Am. (submitted).

Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Stinson, M. R. (**1990**). ''Revision of estimates of acoustic energy reflectance at the human eardrum,'' J. Acoust. Soc. Am. **88**, 1773–78.

Stinson, M. R., Shaw, E. A. G., and Lawton, B. W. (**1982**). ''Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal,'' J. Acoust. Soc. Am. **72**, 766–773.

Voss, S. E., and Allen, J. B. (**1994**). ''Measurement of acoustic impedance and reflectance in the human ear canal,'' J. Acoust. Soc. Am. **95**, 372–384.

Zweig, G., and Shera, C. A. (**1995**). ''The origin of periodicity in the spectrum of evoked otoacoustic emissions,'' J. Acoust. Soc. Am. **98**, 2018–2047.

Zwicker, E. (**1990**). ''On the influence of acoustical probe impedance on evoked otoacoustic emissions,'' Hearing Res. **47**, 185–190.

Basic characteristics of click-evoked otoacoustic emissions in infants and children

Beth A. Prieve and Tracy S. Fitzgerald
Communication Sciences and Disorders, Syracuse University, 805 S. Crouse Avenue, Syracuse, New York 13244-2280

Laura E. Schulte
STAR Consulting, Omaha, Nebraska 68132

(Received 20 June 1996; revised 3 July 1997; accepted 7 July 1997)

Since Kemp [J. Acoust. Soc. Am. **64**, 1386–1391 (1978)] first described click-evoked otoacoustic emissions (COAEs), researchers have advocated their use as an excellent tool for diagnosing hearing loss in infants and children. However, there are few detailed reports of COAEs in this population, and those that do exist suggest that there are age-dependent differences. The purpose of the current study was to determine basic characteristics of COAEs in infants, toddlers, children, and young adults and to define any differences among age groups. An additional goal was to ensure that spontaneous otoacoustic emissions (SOAEs) did not confound any possible developmental effects. COAEs and SOAEs were measured from one ear of 223 normal-hearing subjects. COAE input/output functions indicated that children aged less than one year have higher COAE levels than older children and adults. Children aged 1–5 yr had higher COAE levels than those aged 12–17 yr and adults. These differences were independent of level and SOAE status, but were dependent on frequency. The results of this study suggest that different clinical norms may be necessary for children aged less than 6 years. © *1997 Acoustical Society of America.* [S0001-4966(97)01811-0]

PACS numbers: 43.64.Jb, 43.64.Ri [RDF]

INTRODUCTION

Since the time that Kemp (1978) first described otoacoustic emissions (OAEs), they have been advocated as a sensitive measure of peripheral auditory function. Because they can be measured quickly and noninvasively, many researchers have stated that they are an ideal test for hearing loss in infants and children. However, there are few reports of OAEs in these populations. Existing reports suggest that click-evoked OAEs (COAEs) may be different in neonates, children, and adults. In order for COAEs to be an effective clinical tool for infants and children, it must be determined if there are differences in COAEs with age, and if they exist, to describe what those differences are.

It is clear that neonates (defined as infants less than 1 month old) have significantly higher COAE levels than adults (Collet *et al.*, 1993; Kok *et al.*, 1992, 1993; Norton, 1993) and older children (Norton, 1994). Developmental studies suggest that changes may occur early in a child's life. Engdahl *et al.* (1994) reported that infants aged 3 months and 12 months had lower COAE levels in frequency regions ranging from 2.5 to 4.5 kHz than those aged 3 days. Johnsen *et al.* (1989) studied COAEs measured at birth and then again at 4 years of age in nine subjects. Through observation of single-subject data, they found that COAEs contained less high-frequency energy at 4 years of age than they did at birth.

There are also data to suggest that COAE levels in children are higher than those measured in adults. Spektor *et al.* (1991) reported that COAE levels in five children aged 4–10 years were approximately 4 dB higher than those measured in adults. Norton and Widen (1990) tested COAEs in subjects ranging in age from 17 days to 30 years who were divided into three groups based on age; 0.0–9.9, 10.0–19.9, and 20.0–29.9 yr. They found significant differences in COAE levels evoked by 80 dB peak SPL (pSPL) clicks among all age groups. Bray and Kemp (1987) also found that children aged between 6 and 13 years had higher COAE levels than did adults, but explained the differences were due to smaller ear canal volume in the children.

Although all of these studies provide evidence for changes in COAEs during childhood, the pattern with development is not well defined. Single subject data from Johnsen *et al.* (1989), Norton and Widen (1990), and Prieve (1992) suggest that the most dramatic changes in COAEs may take place between birth and 2–4 yr of age. Therefore data are needed from a larger subject pool grouped into narrower age ranges in order to more accurately detail COAEs in infants, toddlers, preschoolers, and school-age children.

There are even fewer studies that address possible maturation of COAE input/output (I/O) functions in children. In the studies cited above, COAEs were evoked by clicks of approximately 80 dB pSPL (Bray and Kemp, 1987; Engdahl *et al.*, 1994; Kok *et al.*, 1993; Norton and Widen, 1990). It has been shown that neonates have higher COAE levels than adults (Norton, 1993) and children aged 4–13 yr (Norton, 1994) for stimulus levels ranging between 50–80 dB pSPL. However, there are no systematic studies reporting COAE input/output (I/O) in older infants and children.

Additionally, most studies employed a subtraction paradigm when measuring COAEs. A subtraction paradigm was used to eliminate stimulus artifacts that could be mistakenly identified as COAEs, a problem most apparent at high stimulus levels. The subtraction technique hinges on certain as-

sumptions about COAE growth as stimulus level is increased. While the assumptions hold well for high stimulus levels, they are inappropriate for low- and midlevel stimuli (Bray, 1989; Fitzgerald and Prieve, in press; Ravazanni and Grandori, 1993). Because the subtraction procedure makes assumptions about I/O functions, it must be determined whether I/O functions across age groups are similar to ensure that the subtraction paradigm is appropriate for all ages.

Finally, there is the need to address if the presence of spontaneous OAEs (SOAEs) are confounding factors when studying developmental changes of COAEs. COAE I/O functions for adults with SOAEs (Prieve and Falter, 1995; Probst *et al.*, 1986; Zwicker, 1983) are different than for those individuals not having SOAEs. Specifically, individuals having SOAEs have higher COAE levels, lower COAE thresholds, and I/O functions that saturate at lower stimulus levels than individuals not having SOAEs. There is evidence that the prevalence of SOAEs is similar among infants (Burns *et al.*, 1992), children (Strickland *et al.*, 1985), and adults (e.g., Bilger *et al.*, 1990). However, within a particular sample there may be unequal proportions of individuals having SOAEs. Therefore it is important to consider the prevalence of SOAEs in the sample to ensure that any differences among age groups are due to age and not the prevalence of SOAEs within the sample group.

The goal of the current study was to determine if there are significant differences in COAEs evoked by various stimulus levels with age, and if there are differences, to define them. An additional goal was to ensure that SOAE status of the subject was taken into account and to determine if developmental effects are observed in individuals with and without SOAEs.

I. METHODS

COAEs were measured from one ear of 223 subjects aged 4 weeks to 29 years. Data from subjects were placed into one of seven groups based on age: One year or less ($n=35$, $\overline{x}=0.50$, s.d.$=0.22$), 1–3 yr ($n=31$, $\overline{x}=2.56$, s.d. $=0.68$), 4 to 5 yr ($n=29$, $\overline{x}=4.55$, s.d.$=0.55$), 6–8 yr ($n=37$, $\overline{x}=6.95$, s.d.$=0.84$), 9–11 yr ($n=29$, $\overline{x}=9.75$, s.d. $=0.80$), 12–17 yr ($n=32$, $\overline{x}=13.76$, s.d.$=2.20$), and 18–29 yr ($n=30$, $\overline{x}=22.83$, s.d.$=1.73$). Age ranges included in each group were based on single-subject data suggesting that the greatest change in COAEs may occur between birth and 2–4 yr of age, with more gradual changes in older children (Norton and Widen, 1990; Prieve, 1992). Therefore the age range for the youngest group of children was only 1 year, with increasingly wider ranges used for older children. The one exception was for the group aged 1–3 yr. Of the approximately 25 children aged 1.1–1.9 yr who came to participate in the study, only five children were successfully tested. Therefore the data from these five children were grouped with those aged 2 to 3 yr. The group of adults is considered a ''control'' group, as previous research has shown that there are not significant changes in COAEs with adult age once hearing threshold has been taken into consideration (Prieve and Falter, 1995; Stover and Norton, 1993).

Children aged less than 2 years underwent a screening for warble tones presented in a sound field at 15 dB HL *re*: ANSI (1989). This information was obtained on approximately 10% of the children aged less than two years. No audiometric information could be obtained in the remaining 90% of the infants because they were either too young to perform visual reinforcement audiometry or they were uncooperative. Audiograms for subjects aged 2–4 yr were obtained using play audiometry under earphones. If possible, thresholds were determined for at least 500, 1000, 2000, and 4000 Hz. In the worst case, a child was screened at those four frequencies at 15 dB HL. Thresholds for subjects aged 4–29 yr were measured at octave frequencies from 250–8000 Hz and the interoctave frequencies of 3000 and 6000 Hz. Tympanograms using a 226-Hz probe tone were measured before COAE testing in all subjects at the beginning of the test session. Static admittance and equivalent ear canal volume were within published norms (ASHA, 1990). Tympanometric peak pressure ranged from -120 to $+98$ daPa. Mean tympanometric peak pressures among the groups was similar, ranging from -28.06 daPa (for children aged 4 to 5 yr) to -5.17 daPa (children aged 1 year or less).

In addition to the normal-hearing subjects, COAEs were measured on seven individuals with severe-to-profound sensorineural hearing loss. Based on what is known about COAEs from hearing-impaired individuals, no COAEs were expected from the seven subjects, as the best threshold measured was 60 dB HL at any audiometric frequency. However, when a subtraction paradigm is not used, there is often energy present in the COAE response window that is thought to be due to stimulus and middle ear artifact (Bray, 1989; Kemp, 1978; Kemp *et al.*, 1990). The purpose of including these subjects with severe hearing loss was to establish if, or the extent to which, measurement procedures used with the normal-hearing subjects contained response energy not likely to be from nonlinear cochlear processes. Two of the subjects were children, aged 4 and 5 years. The remaining subjects were adults aged 19–78 yr.

COAEs and SOAEs were measured using the ILO88 hardware and software. Stimuli were 80-μs rectangular clicks presented at a rate of 50/s. Stimuli were generated by a 14-bit digital-to-analog converter and presented to the subject via a BP 1712 Knowles receiver housed in a custom-designed probe. Click level was calibrated in the ear canal so that the level was ± 2 dB of 5 dB steps from 40–80 dB pSPL. A miniature microphone (Knowles EA 1843), also housed in the probe, was used to record the stimuli and OAEs. The output of the microphone was amplified, sampled using a 14-bit analog-to-digital converter at a 25-kHz sampling rate and averaged in the time domain.

COAEs were collected using two types of averaging. In the first type, which shall be referred to as ''regular averaging,'' COAEs were evoked by equal-amplitude clicks. In the second type, which shall be referred to as ''subtraction paradigm,'' clicks having different amplitudes are used in order to rid the response of stimulus artifact. It is the ''default'' mode of operation for the ILO88 and the one that is used in most studies. Specifics of the technique have been detailed in previous publications (e.g., Bray, 1989; Fitzgerald and Prieve, in press; Kemp *et al.*, 1990).

Energy in the ear canal was averaged into two, 20-ms

buffers, resulting in two averaged traces (called trace ''A'' and trace ''B''). Traces were blanked from 0 to 2.5 ms to remove any remaining stimulus artifact and then bandpass filtered from 984 to 4882 Hz. To quantify the reproducible portions of the two traces (presumably the COAEs), the two traces were averaged together and the rms amplitude computed. This rms level will be referred to as a ''derived response.'' To estimate the residual noise, one trace (trace ''B'') was subtracted from the other (trace ''A'') and the rms amplitude was computed. COAEs were considered present if the intensity of the reproducible components of the response waveform computed from the fast Fourier transformation (FFT) (cross spectrum powers) exceeded the residual noise estimate. The term ''COAE level'' is used to refer to this measurement.

SOAEs were also measured using the ILO88 system. The technique involves ''time locking'' the SOAE to the ILO88 average by presenting a click stimulus to the ear. The click was presented at a level of 70 dB pSPL at a rate of 12.5/s. Energy present in the ear canal was measured to 80-ms poststimulus time, grouped in 20-ms time blocks. The time waveform occurring at 60–80-ms poststimulus time was analyzed for presence of any SOAEs that had been synchronized to the click. The frequency range of the recording system was 0.012–6250 Hz. The frequency resolution of the FFT was 12.3 Hz. SOAEs measured using clicks to synchronize them produce similar, but not exactly, the same results as the method of placing a microphone in the ear and using free-running averaging. According to Wable and Collet (1994), approximately 96% of SOAEs measured using the synchronizing click and evaluating energy 60–80-ms poststimulus time were also identified as SOAEs using free-running averaging. However, low-amplitude, high-frequency SOAEs measured using free-running averaging sometimes were not detected using the synchronizing click method.

Test procedures were designed to accommodate a wide range of ages and attentional abilities. First, data collection time had to be limited for children aged 4 weeks to 2 years. Second, it was important to measure I/O functions; therefore it was desirable to measure COAEs evoked by low-level clicks. Based on the literature, it was suspected that adults would have low COAE levels, especially those evoked by low-level clicks. In order to maximize capture of low-level COAEs, it was anticipated that more averaging would be necessary. Therefore testing strategy was designed to accommodate short attention spans and detection of COAEs evoked by low-level clicks. Five data points were considered essential: COAEs evoked by click levels of 40, 60, and 80 dB pSPL using regular averaging, those evoked by a click level of 80 dB pSPL using the subtraction paradigm, and SOAEs. The number of averages collected depended on the stimulus level. For the SOAEs and COAEs evoked by clicks of 65–80 dB pSPL, 260 averages were collected and for stimuli ranging between 40 and 60 dB pSPL, 600 averages were collected. In cooperative preschoolers, data were also collected at 50 and 70 dB pSPL. COAEs were collected in 5-dB steps in subjects aged six years and older.

Children 1 year of age and younger, as well as five children aged 1 and 2 years, were sleeping during testing. All other subjects were awake at the time of testing. Noise levels were lower if the child was engaged in a quiet activity. Children aged 4 years and younger were quietly read stories or watched videos with the sound very low. Older children also watched videos.

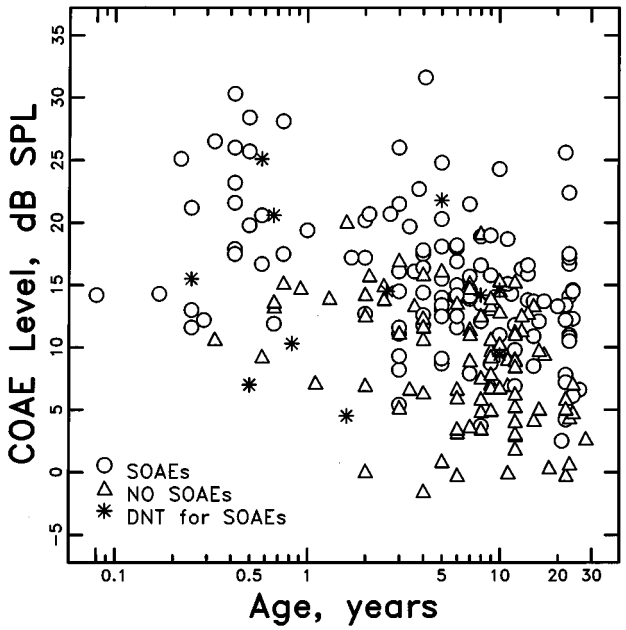


FIG. 1. Scatterplot of COAE level as a function of log age. Open symbols represent data from subjects having measurable SOAEs, open triangles represent data from subjects having no measurable SOAEs and asterisks represent data from subjects in which SOAE testing was not completed.

Off-line analysis: The first 5 ms were eliminated from the COAE traces collected using regular averaging. Based on tests in a 1-cc test cavity, eliminating the first 5 ms of the trace excluded any energy that could have been mistakenly classified as a ''COAE.'' COAEs evoked by 60 and 80 dB pSPL clicks and 80 dB pSPL clicks using the subtraction paradigm were analyzed off-line into 1/3 octave bands by a program written by Dr. Stephen Neely (for details of the program, see Prieve *et al.*, 1993). Only files in which there was a COAE present were analyzed.

II. RESULTS

Figure 1 illustrates COAE level evoked by 80 dB pSPL clicks as a function of subject age. The data were best fit by a log function ($r = 0.41$) indicating that the rate of decrement in COAE level changes more dramatically for children of young ages. Although data analyzed in this way demonstrates that there is a relationship between COAE level and age, it does not probe whether there were significant differences between age groups. The data illustrated in Fig. 1 support the use of groups having narrower age ranges for younger subjects.

Figure 2 shows mean COAE level as a function of click level with the parameter being subject age. Children aged less than 1 year had higher mean COAE levels than all other age groups. Children aged 1–3 yr and 4 to 5 yr had mean COAE levels similar to each other but higher than those of older children. Young adults appeared to have the lowest mean COAE levels. The asterisks indicate response levels

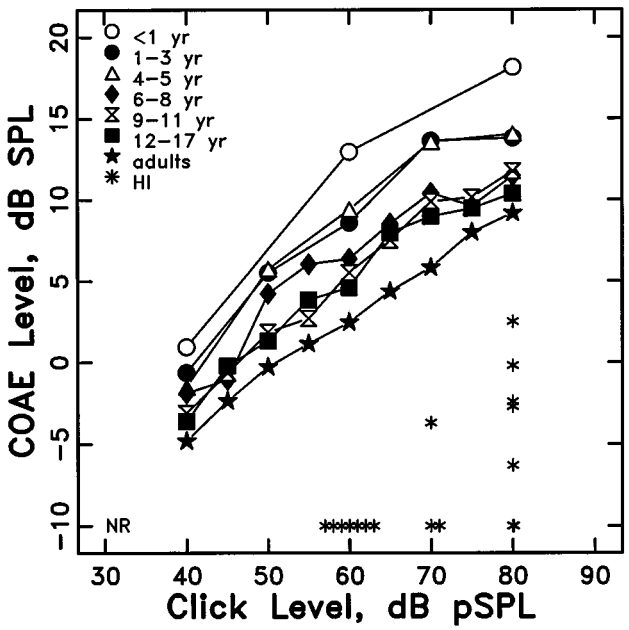


FIG. 2. Mean COAE level as a function of click level for the seven age groups. Each group is represented with a different symbol type. Asterisks represent data from individuals having severe-to-profound sensorineural hearing loss to demonstrate that they do not have COAE levels within the range from normally hearing subjects.

from the severe-to-profound hearing-impaired (HI) individuals. Those data points plotted at a COAE level of −10 dB SPL indicate no response. No subject with hearing loss had ''responses'' at 60 dB pSPL and only one subject with hearing loss had a ''response'' at 70 dB pSPL. There were several hearing-impaired subjects who had ''responses'' to click stimuli of 80 dB pSPL, however, the levels were lower than one standard deviation below the mean for adult subjects. These data are in agreement with Kemp (1978) who found that COAEs evoked by regular averaging produced responses from deaf ears at high stimulus levels. The responses are different than those found in the normal subjects; their levels are low and tend to have a shorter duration (Fitzgerald and Prieve, in press).

Figure 3 shows I/O functions for subjects broken into groups based on SOAE status. I/O functions for those individuals having SOAEs are shown in the left panel, and those without SOAEs are shown in the right panel. Several important trends were apparent. First, mean COAEs were highest in infants and lowest in adults regardless of whether or not SOAEs were present. Second, within each age group, individuals with SOAEs had higher mean COAE levels than those who did not have SOAEs. To test if differences in mean COAE level among age groups and SOAE status were significant, a 7 (group) ×2 (SOAE status) ×2 (stimulus level at 60 and 80 dB pSPL) mixed-model analysis of variance (ANOVA) with repeated measures on the last factor was employed. Results are summarized in Table I. COAE levels were significantly different among age groups ($F_{6,193}=8.52$, $p<0.00005$). These differences did not depend on stimulus level, as there was no significant interaction between stimulus level and group ($F_{1,193}=1.23$, $p=0.2943$). Mean COAE levels were significantly higher for individuals having SOAEs than for those not having SOAEs ($F_{1,193}=88.70$, $p<0.00005$). This was true for all age groups, as there was no significant interaction between group and SOAE status ($F_{6,193}=0.40$, $p=0.8794$), nor was there a significant interaction among stimulus level, group, and SOAE status ($F_{6,193}=0.36$, $p=0.9012$). To determine which groups had significantly higher mean COAE levels than others, a Tukey *post-hoc* test was performed ($\alpha=0.05$ familywise). Children aged less than 1 year had significantly higher mean COAE levels than all other groups. Children aged 1–3 yr and 4 to 5 yr had significantly higher mean COAE levels than teens aged 12–17 yr and young adults. To summarize, there were differences in mean COAE levels among groups. The differences among groups were found for both stimulus levels and regardless of whether the subjects had SOAEs.

TABLE I. Results for 7 (group)×2 (SOAE status)×2 (click level) mixed-model ANOVA with repeated measures on the last factor.

Source	SS	df	MS	*F*	*p*
Between subjects					
Group	2400.78	6	400.13	8.52	0.0000
SOAE	4163.26	1	4163.26	88.70	0.0000
Group*SOAE	112.21	6	8.70	0.40	0.8794
Error	9058.99	193	46.94		
Within subjects					
Level	3131.12	1	3131.12	534.98	0.0000
Level*Group	43.07	6	7.18	1.23	0.2943
Level*SOAE	120.02	1	120.02	20.51	0.0000
Level*Gr*SOAE	12.77	6	2.13	0.36	0.9012
Error	1129.59	193	5.85		

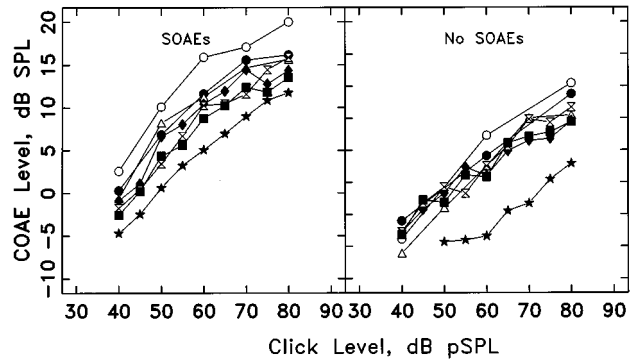


FIG. 3. Input/output functions for individuals having SOAEs (left panel) and individuals not having SOAEs (right panel). Symbols are the same as those in Fig. 2.

There was a significant difference in mean COAE level dependent on stimulus level ($F_{1,193}=534.98$, $p<0.00005$). There was also a significant interaction between stimulus level and SOAE status ($F_{1,193}=20.51$, $p<0.00005$). Follow-up with simple main effects tests indicated that there were significant differences in mean COAE levels evoked by 60 and 80 dB pSPL regardless of SOAE status and that there were significant differences in mean COAE levels evoked by both stimulus levels among individuals having and not having SOAEs. The significant interaction is probably due to the fact that the mean difference in COAE levels evoked by the 60 and 80 dB pSPL clicks for individuals having SOAEs was

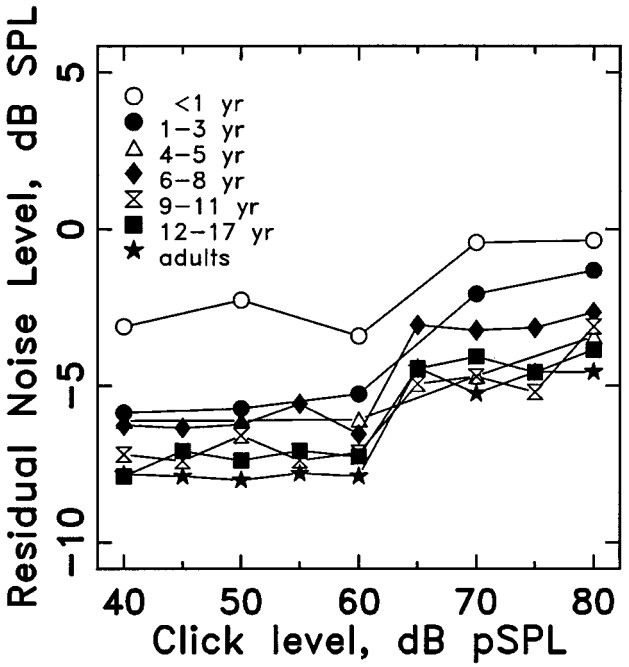


FIG. 4. Mean residual noise levels for the seven age groups. Levels are approximately 3 dB lower for stimuli levels between 35 and 60 dB pSPL than for stimuli 65 dB pSPL and higher because approximately twice as many repetitions were included in the average.

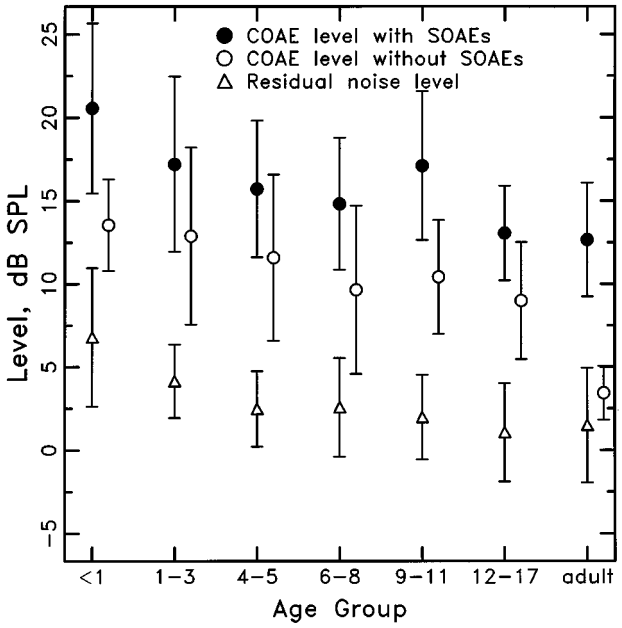


FIG. 5. Mean COAE levels and residual levels for the seven age groups evoked by 80 dB pSPL clicks using the subtraction procedure. Closed circles represent mean COAE levels for individuals having SOAEs and open circles represent mean COAE levels for individuals not having SOAEs. Open triangles represent mean residual levels averaged across all subjects in the age group. Error bars represent ±1 standard deviation of the mean.

only 4.885 dB, whereas the corollary for individuals not having SOAEs was 7.19 dB. Based on other reports in the literature (Prieve and Falter, 1995; Probst *et al.*, 1986; Zwicker, 1983), individuals with SOAEs have I/O functions that saturate at lower stimulus levels, so there would be less of a difference in mean COAE levels between the two stimulus levels. COAEs for individuals not having SOAEs would still not be saturated at 60 dB pSPL, therefore, the difference in mean COAE levels evoked by 60 and 80 dB pSPL clicks would be larger.

Only data from 60 and 80 dB pSPL were included in the three-way ANOVA because COAEs were not detectable in many individuals using 40 dB pSPL clicks. A separate, one-way ANOVA was performed on mean COAE levels evoked by 40 dB pSPL clicks. Mean COAE levels were significantly different among groups ($F_{6,132}=3.54$, $p=0.0028$). A *post-hoc* Tukey test ($\alpha=0.05$ familywise) indicated that children aged less than 1 yr had significantly higher mean COAE levels than teens aged 12–17 yr and adults. There were no other significant differences.

Mean residual noise levels for the seven groups are illustrated in Fig. 4. Noise estimates for click levels ranging between 40 and 60 dB pSPL were approximately 3 dB lower than those for click levels 65–80 dB pSPL, because twice as many averages were collected. Infants aged less than 1 yr had higher noise levels than all other groups. Children aged 1–3 yr had slightly higher noise levels than older children and adults.

Most of the previous literature provides data on COAEs measured using the default mode of the ILO88, which involves collecting COAEs using a subtraction paradigm at a click level of approximately 80 dB pSPL. In order to bridge our data with those in the literature, we also ran this condition on our subjects. The data are illustrated in Fig. 5. The closed symbols represent mean COAE levels for individuals having SOAEs and the open symbols represent mean COAE levels for individuals not having SOAEs. The open triangles represent mean residual noise level for both groups combined. The error bars represent plus and minus one standard deviation from the mean. Similar to COAE levels recorded using regular averaging, infants had the highest mean COAE level, with older children and adults having lower mean COAE levels. Mean COAE levels were lower for individuals not having SOAEs, regardless of age. A one-way ANOVA with group and SOAE status as factors, indicated that there were significant differences in COAE level depending on age ($F_{6,176}=9.98$, $p<0.0005$). COAE levels were higher in individuals having SOAEs than those who did not ($F_{1,176}=71.89$, $p<0.000\,05$) and there was no interaction between group and SOAE status ($F_{6,176}=1.17$, $p=0.3227$). Tukey *post-hoc* analysis ($\alpha=0.05$ familywise) indicated that children aged less than 1 yr had significantly higher COAE levels than those aged 4–17 yr and adults, and children aged 1–5 yr had significantly higher COAE levels than teens aged 12–17 yr and adults.

Because of the differences in COAE levels among groups, it was important to investigate if these differences were frequency dependent. Broadband COAEs and the evoking stimulus were analyzed off-line by a program that parses them into 1/3-octave bands. Among the output provided by the program were the stimulus energy in 1/3-octave bands, rms level of the averaged traces (derived responses), and the residual noise level. Although peak level was calibrated in the ear canal of each subject, it was possible that the spectral content of the click stimuli varied among subject groups. Figure 6 illustrates the mean stimulus level in

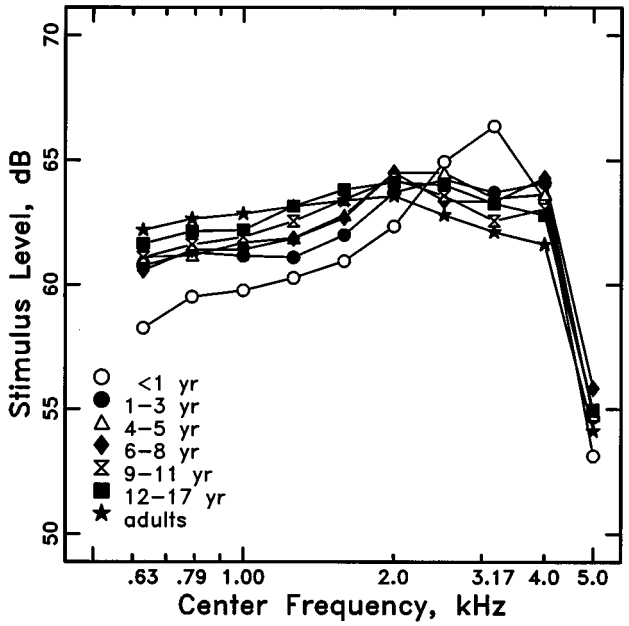


FIG. 6. Stimulus level in dB *re*: equivalent SPL in a 10-kHz bandwidth as a function of 1/3-octave band center frequency. The parameter is age group. The spectral content of the stimulus is different for the children aged less than 1 yr than all other groups.

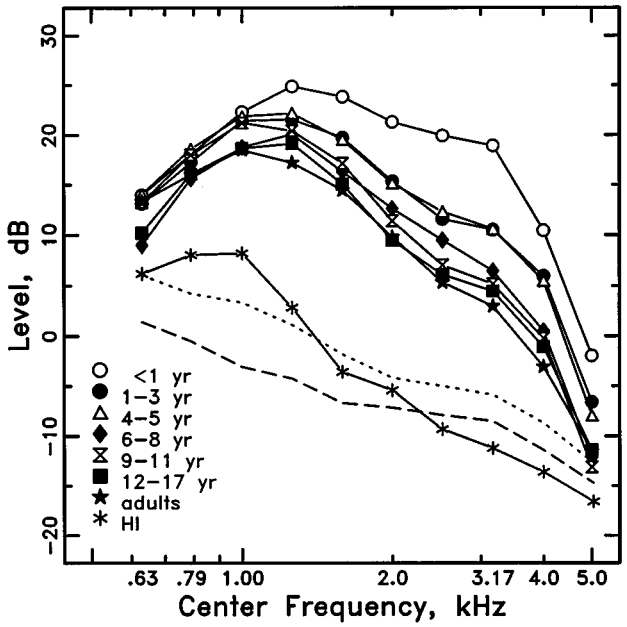


FIG. 7. Mean 1/3-octave band derived response level as a function of center frequency for COAEs evoked by an 80 dB pSPL click collected using regular averaging. Level is equivalent to dB SPL in a bandwidth of 10 kHz. The dotted line represents mean residual level in children aged less than 1 yr and the dashed line represents the corollary for adults. The function represented by asterisks is the mean-derived response level for seven individuals having severe-to-profound sensorineural hearing loss.

1/3-octave bands for clicks calibrated to 80 dB pSPL. It can be seen that the mean spectral content of the click measured in the infant ear canal was different than that delivered to older children and adults. One-third-octave band levels were weaker for bands centered from 630 to 2000 Hz and higher for bands centered at 2520 and 3170 Hz than in the other subjects. To determine whether differences in stimulus levels among groups were significant, a 7 (group) $\times 10$ (frequency) mixed-model ANOVA with repeated measures on the latter factor was performed. There was a significant group by frequency interaction ($F_{54,1719}=4.68$, $p<0.00005$) but not a significant group main effect ($F_{6,191}=1.51$, $p=0.1759$) indicating that there were significant differences among groups for particular frequency bands. Simple main effects tests and *post-hoc* Tukey tests ($\alpha=0.05$ familywise) revealed that children aged less than 1 yr had significantly lower stimulus levels in the 630-, 790-, 1000-, 1260-, and 1590-Hz bands than teens aged 12–17 yr and young adults. They had significantly higher stimulus levels than all other groups in the 3170-Hz band and significantly higher stimulus levels in the 2520-Hz band than did adults. There were no significant differences in stimulus levels in the three remaining 1/3-octave bands (2000, 4000, and 5040).

Figure 7 illustrates the mean 1/3-octave-band derived response levels evoked by 80 dB pSPL clicks in each of the seven groups. COAEs were collected using regular averaging. It can be seen that children aged less than 1 yr had higher mean-derived response levels in 1/3-octave bands centered at frequencies of 1260 Hz and higher. Mean-derived response levels in bands centered at frequencies from 1590 to 5040 Hz in children aged 1–3 yr and 4 to 5 yr were lower in level than those from children aged less than 1 yr, but higher than older children and adults. The spectral contents of the derived responses for all other groups were similar to each other. To investigate if these differences were statistically significant, a 7 (group) $\times 10$ (frequency) mixed-model ANOVA with repeated measures on the last factor was performed. There was a significant frequency by group interaction ($F_{54,1980}=5.21$, $p<0.00005$) indicating that there were significant differences in mean-derived response levels across groups for particular frequencies. To follow-up, simple main effects tests were run for each band. There were significant differences in the mean-derived response level across group in all 1/3-octave bands except for those centered at 790 and 1000 Hz. To ensure that differences were due to age and not to differences in the spectral content of the stimuli, analyses of covariance (ANCOVAs) were attempted using the stimulus level in each 1/3-octave band as the covariate. ANCOVA analysis was appropriate for two frequencies bands, 2520 and 3170 Hz. ANCOVAs were also significant, indicating that there were differences among age groups when controlling for stimulus levels. *Post-hoc* Tukey tests ($\alpha=0.05$ familywise) were performed and the results are summarized in the second column of Table II. Children aged less than 1 yr had significantly higher mean-derived response levels than most other groups for bands centered at frequencies from 1260 to 5040 Hz. Children aged 1–3 yr had higher mean-derived response levels than teens and adults for bands centered at 1590, 2000, and 4000 Hz and higher levels than adults for the 3170-Hz band. For the 5040-Hz band, children aged 1–3 had higher mean-derived response levels than children aged 6–11 yr. Children aged 4 to 5 yr had higher mean-derived response levels than teens and adults in the 2000-, 2529-, and 4000-Hz bands and higher than adults in the 1260-, 1590-, and 3170-Hz bands.

The dotted line in Fig. 7 represents mean residual noise

TABLE II. Significant differences in mean-derived response levels among groups for 1/3-octave bands. COAEs were evoked by 80 dB pSPL clicks and were recorded using regular averaging. *Post-hoc* Tukey tests ($\alpha=0.05$ familywise) were performed for frequency bands having significant simple main effects (630, 1000, 1590, 2000, 4000, and 5040 Hz). ANCOVAs were performed for 1/3-octave bands centered at 2529 and 3170 Hz using the stimulus level in that band as the covariate.

Center frequency, Hz	80 dB pSPL Significant differences	60 dB pSPL Significant differences
630	<1 yr, 4 to 5 yr>6–8 yr	none
790	none	none
1000	none	none
1260	<1 yr>4–17 yr, adults 4 to 5 yr>adults	<1 yr>adults 4 to 5 yr>adults
1590	<1 yr>4–17 yr, adults 1–3 yr>12–17 yr, adults 4 to 5 yr>adults	<1 yr>12–17 yr, adults 4 to 5 yr>adults
2000	<1 yr>1–17 yr, adults 1–5 yr>12–17 yr, adults	<1 yr>6–17 yr, adults
2520	<1 yr>1–17 yr, adults 4 to 5 yr>12–17 yr, adults	<1 yr>1 to 3 yr, 6–17 yr, adults 4 to 5 yr>adults
3170	<1 yr>4–17 yr, adults 1–5 yr>adults	<1 yr>1–17 yr, adults 1–5 yr>adults
4000	<1 yr>4–17 yr, adults 1–5 yr>12–17 yr, adults	<1 yr>6–17 yr, adults 1–3 yr>adults 4 to 5 yr>12–17 yr, adults
5040	<1 yr>4–17 yr, adults 1–3 yr>6–11 yr	<1 yr>6–17 yr, adults

level for children aged less than 1 yr and the dashed line represents the corollary for adults. Children aged less than 1 yr had higher mean residual noise levels than adults at all frequencies, but not equally so. Mean noise levels in higher-frequency bands between the two groups were more similar than those in lower frequency bands. The function depicted by asterisks represents mean-derived response level in the seven hearing-impaired ears. These data illustrate that using regular averaging there was some low-frequency energy present, presumably due to passive mechanisms. One can only assume that this low-frequency energy was also present in the COAEs measured from the normal-hearing ears. However, it is much lower in level, and although overall energy in the band reflects both, it is probable that the greatest contribution to COAE level in lower frequency bands from normal-hearing ears was from active mechanisms.

Derived response levels evoked by 60 dB pSPL clicks using regular averaging were also analyzed into 1/3-octave bands. The shapes of the functions were similar to those shown for the 80 dB pSPL stimuli, so a figure was not included. Similar to COAEs evoked by 80 dB pSPL clicks, a 7(group)×10(frequency) mixed-model ANOVA with repeated measures on the last factor indicated a significant frequency by group interaction ($F_{54,1935}=4.75, p<0.00005$). Simple main effects tests were run for every 1/3-octave frequency except for bands centered at 2520 and 3170 Hz, where ANCOVAs were appropriate. *Post-hoc* Tukey tests ($\alpha=0.05$) were performed and the results listed in the last column of Table II. The results are similar to, but not exactly the same as the significant differences found among groups using the 80 dB pSPL click.

Figure 8 is similar to Fig. 7 except that it represents

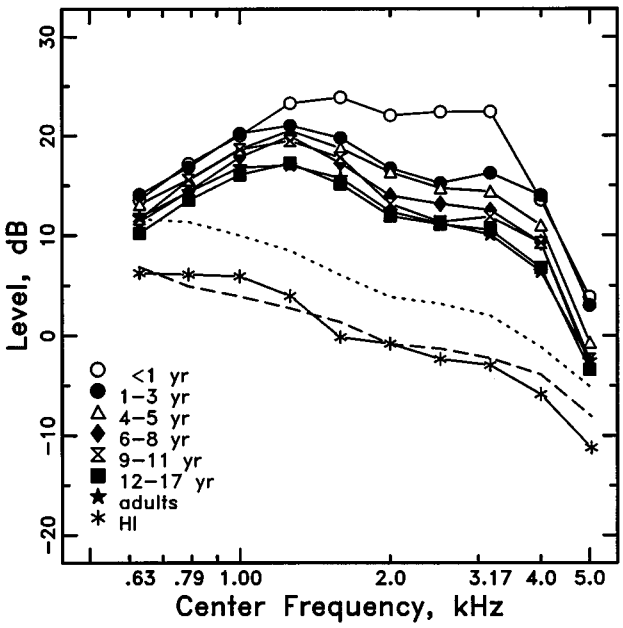


FIG. 8. Similar to Fig. 6, but for COAEs evoked by 80 dB pSPL clicks using the subtraction method.

1/3-octave-band-derived response levels evoked by the 80 dB pSPL click using the subtraction procedure. Several points are apparent. First, energy in 1/3-octave bands ranging from 2520 to 5040 Hz had higher levels than those collected using regular averaging. For regular averaging, the first 5 ms of the trace were eliminated in order to reduce stimulus artifact, also resulting in reduction of short-latency, high-frequency COAEs. Second, the noise estimate was approximately 6 dB higher using the subtraction procedure. This was due to differences in the averaging procedures that have been discussed elsewhere (Fitzgerald and Prieve, in press; Ravazanni and Grandori, 1992). Third, mean-derived response levels from hearing-impaired individuals were lower in the 790- and 1000-Hz bands, but only slightly so. The responses that grew linearly with stimulus level in these individuals had been reduced due to the subtraction procedure. Additionally, the noise level was higher so even if COAEs were the same amplitude, they would be less likely to be classified as a COAE because of the low signal-to-noise ratio. Last, the pattern of derived response level across frequency bands for the different age groups was similar to that seen for derived responses measured using regular averaging. To investigate if the differences were statistically significant, a 7 (group)×10 (frequency) mixed-model ANOVA with repeated measures on the last factor was performed. Similar to results for derived responses using regular averaging, there was a significant frequency by group interaction ($F_{54,1980}=5.21, p<0.00005$) indicating that there were significant differences in derived response level across groups for particular frequencies. Simple main effects tests were run at each frequency. ANCOVAs were appropriate for the two frequency bands centered at 2529 and 3170 Hz. There were no significant differences across groups for bands centered at 630, 790, and 1000 Hz. Simple main effects tests were significant at all other frequencies. *Post-hoc* Tukey tests ($\alpha=0.05$ familywise) were performed for all other frequency

TABLE III. Significant differences in mean-derived response levels among groups for 1/3-octave bands. COAEs were evoked by 80 dB pSPL clicks and were recorded using the subtraction method. *Post-hoc* Tukey tests ($\alpha = 0.05$ familywise) were performed for frequency bands having significant simple main effects (630, 1000, 1590, 2000, 4000, and 5040 Hz). ANCOVAs were performed for 1/3-octave bands centered at 2529 and 3170 Hz using the stimulus level in that band as the covariate.

Center frequency, Hz	80 dB pSPL, nonlinear averaging Significant differences
630	none
790	none
1000	none
1260	<1 yr>12–17 yr, adults
1590	<1 yr>4–17 yr, adults
2000	<1 yr>1–17 yr, adults
2520	<1 yr>1–17 yr, adults
3170	<1 yr>1–17 yr, adults
	1–3 yr>12–17 yr
4000	<1 yr>4–17 yr, adults
	1–3 yr>12–17 yr, adults
5040	<1 yr>4–17 yr, adults
	1–3 yr>4–17 yr

bands and the results are summarized in Table III. Children aged less than 1 yr had significantly higher mean-derived response levels than most other groups for bands centered at frequencies ranging from 1260 to 5040 Hz. Children aged 1–3 yr had higher mean-derived response levels than teens in the 3170-Hz band and higher than teens and adults in the 4000- and 5040-Hz bands.

III. DISCUSSION

The main finding of the study is that there were significant differences in mean COAE levels among children of different ages and adults. For COAEs collected using regular averaging, children aged less than 1 yr had higher COAE levels than toddlers, children, teens, and young adults. Children aged 1–5 yr had higher COAE levels than teens and young adults. These changes were frequency dependent, but were independent of stimulus level and SOAE status.

An important aspect of the results is that COAE levels were different among age groups for three stimulus levels. Previous reports indicated that neonates had significantly higher COAE levels for click levels from 35 to 80 dB pSPL than did children aged 4–13 yr (Norton, 1994) and adults (Norton, 1993), the latter differences undergoing statistical testing. The data from the current study indicated that COAE levels measured in infants aged 4 weeks to 1 yr were significantly higher than those measured in older children and adults for low-(40 dB pSPL), mid-(60 dB pSPL), and high-(80 dB pSPL) level clicks. In addition, COAE levels evoked by 60 and 80 dB pSPL clicks were higher in children aged 1–5 yr than in teens and adults.

Although the slopes of the I/O functions were not computed and compared across groups, the lack of a group×level interaction suggests that the shapes of the I/O functions for mid- and high-stimulus levels were similar across age groups. These results suggest that basic properties of the COAE generators are functional even in young infants and that they do not change with age. However, whatever is responsible for absolute COAE level differs in infants and young children. For the clinician, the similarity of the shape of the I/O functions across groups supports the notion that the subtraction method for acquiring COAEs is appropriate for use with these age groups. Additionally, the data imply that there is not ''one'' stimulus level for which the same COAE norms could be used for all ages.

COAEs evoked by clicks presented at 80 dB pSPL using the subtraction method were also significantly different among age groups. Children aged less than 1 yr had higher COAE levels than those aged 4–17 yr and adults. Children aged 1–5 yr had higher COAE levels than those aged 12–17 yr and adults. Results from the current study cannot be directly compared to other published studies because it included infants rather than neonates, and children were grouped into narrower age ranges (Bray and Kemp, 1987; Collet *et al.*, 1993; Kok *et al.*, 1992, 1993; Norton and Widen, 1990; Spektor *et al.*, 1991). The current study is most similar to that reported by Norton and Widen (1990) who found that subjects aged 0.0–9.9, 10.0–19.9, and 20–29.9 yr had significantly different COAE levels. The findings from the current study are similar in some respects. In the current study, children aged less than 1 yr and those aged 1–5 yr had higher COAE levels than children aged 12–17 yr and adults. Although Norton and Widen (1990) grouped all these ages together, they reached similar conclusions. However, there were no significant differences among children aged 9–11 yr or 12–17 yr and adults in the current study, whereas Norton and Widen (1990) found significantly higher COAEs in teens aged 10–19.9 yr than adults aged 20–20.9 yr. There are two differences that might account for the varying results. Norton and Widen (1990) reported the number of ears and do not specify if both ears from the same subjects were included in the statistical tests. In the current study, only one ear per subject was tested. Including statistics from two ears could reduce variability because COAE level between the two ears are significantly correlated (e.g., Robinette, 1992). Also, Norton and Widen (1990) did not report prevalence of SOAEs in each group. It is possible that the group aged 10–19 yr had a higher prevalence of SOAEs than did the adult group, confounding if the significant level difference between the two groups was due solely to age. In terms of absolute levels, the COAE levels for children in the current study are similar to those reported by others (Nozza and Sabo, 1992; Welzl-Muller and Stephan, 1994). COAE levels for children aged less than 1 yr are similar to those reported by Engdahl *et al.* (1994).

The spectrum of the click was different for children aged less than 1 yr and adults, even though absolute click levels were not significantly different. Similar findings have been reported by others (Bergman *et al.*, 1995; Norton, 1993; Spektor *et al.*, 1991). The investigators of these studies thought that the differences in spectral content were mainly due to the different probe design used for neonates. Because of this speculation, the adult probe was used in infant ears in the current study. Despite the fact that the same probe was used for all subjects regardless of age, the spectrum of the click in the infant ear had more energy around 3000 Hz and less energy for frequency bands from 630 to 1000 Hz than in

other age groups. In fact, the click spectrum for the infants looks very similar to that reported by Bergman *et al.* (1995) for click stimuli in neonates using the neonatal probe. The spectral content of the click, especially the lower levels in the low-frequency bands, could be due to energy leakage out of the ear canal.

Another important finding of the current study is that changes in derived response levels were frequency dependent. Previous reports have shown differences in mid- to high-frequency energy between neonates and adults (Spektor *et al.*, 1991; Norton, 1993) and between neonates and a group of older children and adults (Bergman *et al.*, 1995). The levels obtained in each band for the infants in the current study were similar to those reported by Bergman *et al.* (1995) for graduates of the NICU. Other reports suggest that frequency content of COAEs change with age, but such an effect was not rigorously studied (Norton and Widen, 1990; Prieve, 1992; Spektor *et al.*, 1991). The results from the current study indicate that derived response levels in 1/3-octave bands centered from 1260 to 5040 Hz were significantly higher in children aged less than 1 yr than in older children and adults. Children aged 1–5 yr had higher derived response levels in some of those bands than did teens and adults. Derived response levels for bands centered at 1000 Hz and lower generally did not differ among groups. It is possible that the high-pass filtering cutoff of 984 Hz could have affected these comparisons. These results suggest that the differences in overall COAE level across age are at least partially due to reduction of COAE energy in mid- to high-frequency bands.

There were slight variations in which 1/3-octave-band derived response levels were significantly different when the two stimulus levels and different recording paradigms were used. Differences among groups for derived response levels evoked by click levels of 60 and 80 dB pSPL were similar except for children aged 1–3 yr at 1590 and 5040 Hz and those aged 1–5 yr at 2000 Hz. At 5040 Hz, the derived response level in the band was essentially the same as the noise floor for all subject groups except the infants. It is not clear what accounts for the differences in the 1590- and 2000-Hz bands. The main difference in derived response levels evoked by the two stimulus levels (besides the obvious that COAE levels evoked by 60 dB pSPL clicks are lower than those evoked by 80 dB pSPL clicks) is that there is virtually no artifact from ears with hearing loss for click levels at 60 dB pSPL. However, it is unlikely that artifact is contributing to the noted differences with age because hearing-impaired ears do not have significant energy in the frequency bands of 1590 and 2000 Hz at 80 dB pSPL. It also appears that the subtraction technique and regular averaging vary slightly in the significant differences among groups for clicks presented at 80 dB pSPL. The most noticeable disparity is that children aged 4 to 5 yr had no significant differences from adults in any frequency band, yet their overall level analyzed via a one-way ANOVA was significantly higher.

Another major finding of the present study was that individuals having SOAEs had higher COAE levels than individuals not having SOAEs. This has been shown previously for adults (Kulawiec and Orlando, 1995; Prieve and Falter, 1995; Probst *et al.*, 1986; Zwicker, 1983). Results from the current study indicate this finding is true for children as well. Although age was a significant factor accounting for different COAE levels among groups, it accounted for less of the variance ($\eta^2 = 11.9\%$) than did SOAE status ($\eta^2 = 20.6\%$). Prieve and Falter (1995), who studied COAE level in a group of young (19–29 yr) and middle-age adults (40–61 yr), also found that SOAEs accounted for more variance in COAE level than did age. Adult age accounted for 4.3% of the variance and SOAE status accounted for 40.23% of the variance in that study. In the sample of children from the current study, the variance in COAE level accounted for by age was higher and the variance in COAE level accounted for by SOAE status was lower than that reported for the sample consisting of only adults. These findings suggest that, besides hearing threshold, SOAE status influences COAE level across a wide range of ages, while the factor of age is limited to infants and children.

Although the results from the current study indicate that COAEs change with development, two minor points must be mentioned. First, there was a difference in state of awareness among groups. All infants were tested during natural sleep whereas testing for almost all other subjects was completed while they were awake. Data indicate that COAE levels in adults increase during the first 4 h of sleep (Froelich *et al.*, 1993b). and, therefore, cannot be ruled out that some of the increased COAE level in infants was due to sleep. However, state of awareness is probably not the sole source of the higher COAE levels in infants. This conclusion is supported by the finding that children aged 1–5 yr also had significantly higher COAE levels than older subjects, and almost all subjects in these groups were awake during testing. Second, a possible source of variability was introduced in children because toddlers, preschoolers, and young children usually watched videos or listened to books during testing. Previous reports have shown statistically lower COAE levels during visual and auditory attention. Mean differences are typically 0.5 dB, and the subjects in these experiments were told to attend to either visual (Froehlich *et al.*, 1993a) and/or visual and auditory stimuli (Froehlich *et al.*, 1993a; Giard *et al.*, 1994; Meric and Collet, 1993). Again, it cannot be ruled out that COAE levels for young children who watched videos or listened to stories were reduced in level, however, the reported differences are so small that it is unlikely that attention had a major impact on COAE levels in these groups.

It is uncertain why COAE levels are different among age groups, but there are several factors to consider. The first factor is growth of the ear canal. The estimated ear-canal diameter and length increase dramatically through the first year of life, and even at 12 months of age, are less than that of adults (Keefe *et al.*, 1993). It is possible that other anatomical differences between adult ear canals and infant ear canals, such as tissue composition and angle of the tympanic membrane with respect to the ear canal, could also affect OAEs (Northern and Downs, 1991). A second factor that could account for differences in COAE levels between children and adults is the middle ear. Middle ear volume is

smaller in neonates, and increases with development (Anson and Donaldson, 1981). Studies of acoustic impedance, as measured indirectly through tympanometry (Himmelfarb *et al.*, 1987; Holte *et al.*, 1991) and by acoustic reflectance using a broadband stimulus (Keefe *et al.*, 1993) also indicate developmental changes. It is likely that developmental changes in OAEs are due to external and middle ear maturation because of evidence suggesting that the cochleae in neonates aged approximately 40 conceptional weeks are mature (Abdala, 1996; Abdala *et al.*, 1996; Bargones and Burns, 1988; Gorga *et al.*, 1989).

The finding that there are developmental differences in COAEs has clinical implications. First, separate norms may be necessary for children under the age of 6 yr, regardless of the stimulus level chosen. Ultimately, clinical decisions should be based on how well COAEs separate normal and hearing-impaired ears. Previous reports indicate that COAE level, signal-to-noise ratio, and percent reproducibility between two COAE traces are effective for identifying hearing-impaired ears in mixed populations of adults and children (Gorga *et al.*, 1993; Hurley and Musiek, 1994; Prieve *et al.*, 1993). Similar reports suggest the same for children (Glattke *et al.*, 1995; Norton, 1994). However, no study exists in which identification of hearing loss in infants and children using COAEs is compared to identification of hearing loss in adults using COAEs. Therefore it is uncertain whether different criteria will be necessary for the two populations.

ACKNOWLEDGMENTS

Thanks are given to Lara Allesandrelli, Patti Baran Kopylczak, Heather Grossman, Jennifer Horn, Holly Huta, Ann Jackson, Tara McKinney, Nadia Michail, and Deb Rotman Dauber for their assistance with data collection and analysis. Dr. Stephen T. Neely provided the software to parse the broadband stimuli and COAEs into 1/3-octave bands. Dr. Karen Doherty and Dr. Ted Glattke and an anonymous reviewer provided helpful comments on an earlier version of this manuscript. This work was funded by the Deafness Research Foundation and by NIH Grant No. 5R29 DC02028.

Abdala, C. (**1996**). ''DPOAE amplitude as a function of $f2/f1$ frequency ratio and primary tone level separation in human adults and neonates,'' J. Acoust. Soc. Am. **100**, 3726–3740.

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

Anson, B. J., and Donaldson, J. A. (**1981**). *Surgical Anatomy of the Temporal Bone and Ear* (Saunders, Philadelphia).

ANSI (**1989**). ANSI S3.6-1989, ''Specification for audiometers'' (American National Standards Institute, New York).

American Speech-Language Hearing Association (**1990**). ''Guidelines for screening for hearing impairments and middle ear disorders,'' ASHA **32** (Suppl. 2), 17–24.

Bargones, J., and Burns, E. (**1988**). ''Suppression tuning curves for spontaneous otoacoustic emissions in infants and adults,'' J. Acoust. Soc. Am. **83**, 1809–1816.

Bergman, B. M., Gorga, M. P., Neely, S. T., Kaminski, J. R., Beauchaine, K. L., and Peters, J. (**1995**). ''Preliminary descriptions of transient-evoked and distortion-product otoacoustic emissions from graduates of an intensive care nursery,'' J. Am. Acad. Audiol. **6**, 150–162.

Bilger, R. C., Matthies, M. L., Hammel, D. R., and Demorest, M. E. (**1990**). ''Genetic implications of gender differences in the prevalence of spontaneous otoacoustic emissions,'' J. Speech Hear. Res. **33**, 418–432.

Bray, P. J. (**1989**). ''Click evoked otoacoustic emissions and the development of a clinical otoacoustic hearing test instrument,'' Ph.D. dissertation, London University.

Bray, P., and Kemp, D. (**1987**). ''An advanced cochlear echo technique suitable for infant screening,'' Br. J. Audiol. **21**, 191–204.

Burns, E. M., Arehart, K. H., and Campbell, S. L. (**1992**). ''Prevalence of spontaneous otoacoustic emissions in neonates,'' J. Acoust. Soc. Am. **91**, 1571–1575.

Collet, L., Gartner, M., Veuillet, E., Moulin, A., and Morgon, A. (**1993**). ''Evoked and spontaneous otoacoustic emissions: A comparison of neonates and adults,'' Brain Develop. **15**, 249–252.

Engdahl, B., Arnesen, A. R., and Mair, I. W. S. (**1994**). ''Otoacoustic emissions in the first year of life,'' Scand. Audiol. **23**, 195–200.

Fitzgerald, T. S., and Prieve, B. A. (**1997**). ''COAE thresholds: I. Effects of equal-amplitude vs. subtraction methods,'' J. Speech Hear. Res. **40**, 1164–1176.

Froehlich, P., Collet, L., and Morgon, A. (**1993a**). ''Transiently evoked otoacoustic emission amplitudes change with changes of directed attention,'' Physiol. Behav. **53**, 679–682.

Froehlich, P., Collet, L., Valatx, J. L., and Morgon, A. (**1993b**). ''Sleep and active cochlear micromechanical properties in human subjects,'' Hearing Res. **66**, 1–7.

Giard, M.-H., Collet, L., Bouchet, P., and Pernier, J. (**1994**). ''Auditory selective attention in the human cochlea,'' Brain Res. **633**, 353–356.

Glattke, T. J., Pafatis, I. A., Cumminskey, C., and Herer, G. (**1995**). ''Identification of hearing loss in children and young adults using measures of transient otoacoustic emission reproducibility,'' Am. J. Audiol. **4**, 67–82.

Gorga, M. P., Kaminski, J., Beauchaine, K., Jestead, W., and Neely, S. (**1989**). ''Auditory brainstem responses from children 3 months to 3 years of age: Normal patterns of response,'' J. Speech Hear. Res. **32**, 281–288.

Gorga, M. P., Neely, S. T., Bergman, B. M., Beauchaine, K. L., Kaminski, J. R., Peters, J., Schulte, L., and Jesteadt, W. (**1993**). ''A comparison of transient-evoked and distortion product otoacoustic emissions in normal-hearing and hearing-impaired subjects,'' J. Acoust. Soc. Am. **94**, 2639–2648.

Himmelfarb, M. Z., Popelka, G. R., and Shannon, E. (**1987**). ''Tympanometry in normal neonates,'' J. Speech Hear. Res. **22**, 179–191.

Holte, L., Margolis, R., and Cavanagh, R. (**1991**). ''Developmental changes in multifrequency tympanograms,'' Audiology **30**, 1–24.

Hurley, R., and Musiek, F. (**1994**). ''Effectiveness of transient-evoked otoacoustic emissions (TEOAEs) in predicting hearing level,'' J. Am. Acad. Audiol. **5**, 195–203.

Johnsen, N. J., Parbo, J., and Elberling, C. (**1989**). ''Evoked acoustic emissions from the human ear: V: Developmental changes,'' Scand. Audiol. **18**, 59–62.

Keefe, D. H., Bulen, J. C., Arehart, K. H., and Burns, E. M. (**1993**). ''Ear-canal impedance and reflection coefficient in human infants and adults,'' J. Acoust. Soc. Am. **94**, 2617–2638.

Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. T., Ryan, S., and Bray, P. (**1990**). ''A guide to the effective use of otoacoustic emissions,'' Ear Hear. **11**, 93–105.

Kok, M. R., van Zanten, G. A., and Brocaar, M. P. (**1992**). ''Growth of evoked otoacoustic emissions during the first days postpartum,'' Audiology **31**, 140–149.

Kok, M. R., van Zanten, G. A., Brocaar, M. P., and Wallenburg, H. C. S. (**1993**). Click-evoked oto-acoustic emissions in 1036 ears of healthy newborns,'' Audiology **21**, 213–224.

Kulawiec, J. T., and Orlando, M. S. (**1995**). ''Contribution of spontaneous otoacoustic emissions to the click evoked otoacoustic emissions,'' Ear Hear. **16**, 515–520.

Meric, C., and Collet, L. (**1993**). ''Comparative influence of repeated measurement and of attention on evoked otoacoustic emissions,'' Acta Oto-Laryngol. **113**, 471–477.

Northern, J. L., and Downs, M. P. (**1991**). *Hearing in Children* (Williams and Wilkins, Baltimore), 4th ed.

Norton, S. J. (**1993**). ''Application of transient evoked otoacoustic emissions to pediatric populations,'' Ear Hear. **14**, 65–73.

Norton, S. J. (**1994**). '' Emerging role of evoked otoacoustic emissions in

neonatal hearing screening,'' Am. J. Otology **15** (Suppl. 1) 4–12.
Norton, S. J., and Widen, J. E. (**1990**). ''Evoked otoacoustic emissions in normal-hearing infants and children: Emerging data and issues,'' Ear Hear. **11**, 121–127.
Nozza, R. J., and Sabo, D. L. (**1992**). ''Transiently evoked OAE for screening school-age children,'' Hearing J. **45**(11), 29–31.
Prieve, B. A. (**1992**). '' Otoacoustic emissions in infants and children: Basic characteristics and clinical application,'' Seminars Hear. **13**, 37–52.
Prieve, B. A., and Falter, S. R. (**1995**). ''COAEs and SSOAEs in adults with increased age,'' Ear Hear. **16**, 521–528.
Prieve, B. A., Gorga, M. P., Schmidt, A. S., Schulte, L., Peters, J., and Jesteadt, W. (**1993**). ''Analysis of transient-evoked otoacoustic emissions in normal and hearing-impaired ears,'' J. Acoust. Soc. Am. **93**, 3308–3319.
Probst, R., Lonsbury-Martin, B. L., Martin, G. K., and Coats, A. C. (**1986**). ''Otoacoustic emissions in ears with hearing loss,'' Am. J. Otolaryngol. **8**, 73–81.
Ravazanni, P., and Grandori, F. (**1993**). ''Evoked otoacoustic emissions: Nonlinearities and response interpretation,'' IEEE Trans. Biomed. Eng. **40**, 500–505.
Robinette, M. S. (**1992**). ''Clinical observations with transient evoked otoacoustic emissions with adults,'' Seminars Hear. **13**, 23–36.
Spektor, A., Leonard, G., Kim, D. O., Jung, M. D., and Smurzynski, J. (**1991**). ''Otoacoustic emissions in normal and hearing-impaired children and normal adults,'' Laryngoscope **101**, 965–976.
Stover, L., and Norton, S. (**1993**). ''The effects of aging on otoacoustic emissions,'' J. Acoust. Soc. Am. **94**, 2670–2681.
Strickland, E. A., Burns, E. M., and Tubis, A. (**1985**). ''Incidence of spontaneous otoacoustic emissions in children and infants,'' J. Acoust. Soc. Am. **78**, 931–935.
Wable, J., and Collet, L. (**1994**). ''Can synchronized otoacoustic emissions really be attributed to SOAEs?'' Hearing Res. **80**, 141–145.
Welzl-Muller, K., and Stephan, K. (**1994**). ''Confirmation of transiently evoked otoacoustic emissions based on user-independent criteria,'' Audiology **33**, 28–36.
Zwicker, E. (**1983**). ''Delayed evoked oto-acoustic emissions and their suppression by Gaussian-shaped pressure impulses,'' Hearing Res. **11**, 359–371.

Basic characteristics of distortion product otoacoustic emissions in infants and children

Beth A. Prieve and Tracy S. Fitzgerald
Communication Sciences and Disorders, Syracuse University, 805 S. Crouse Avenue, Syracuse, New York 13244-2280

Laura E. Schulte
STAR Consulting, Omaha, Nebraska 68132

David T. Kemp
Institute of Laryngology and Otology, University College London Medical School, 330/332 Gray's Inn Road, London WC1X 8EE, England

(Received 20 June 1996; revised 3 July 1997; accepted 7 July 1997)

Distortion-product otoacoustic emissions at the $2f1-f2$ frequency (DPOAEs) are being advocated as a clinical tool for diagnosis of peripheral auditory pathology. Because they can be measured quickly and noninvasively, they may be an excellent method for identifying hearing loss in infants and children. However, few studies have examined the characteristics of DPOAEs in infants and children or detailed if, and how, their responses differ from those of adults. The purpose of the current study was to determine basic characteristics of DPOAEs in infants, toddlers, children, and young adults and to define any differences among age groups. An additional goal was to ensure that the presence of spontaneous otoacoustic emissions (SOAEs) did not confound any developmental effect. DPOAE input/output (I/O) functions at seven $f2$ frequencies and SOAEs were measured from one ear of 196 subjects. Children aged less than 1 yr had significantly higher mean DPOAE levels than older children and adults, and children aged 1–3 yr had higher mean DPOAE levels than teens and adults. These differences were dependent on frequency but were independent of $f2$ level and SOAE status. At every $f2$ frequency, groups of individuals having SOAEs had higher mean DPOAE levels than those not having SOAEs. [S0001-4966(97)01911-5]

PACS numbers: 43.64.Jb, 43.64.Ri [RDF]

INTRODUCTION

Using evoked otoacoustic emissions (EOAEs) as a clinical tool for diagnosis of hearing loss has gained acceptance in the past few years. Predominantly two types of EOAEs are being considered for clinical use; click-evoked OAEs (COAEs) and distortion-product OAEs occurring at the $2f1-f2$ frequency (DPOAEs). Although the majority of information about DPOAEs has been obtained from adults, they may prove to be most useful in the pediatric population by providing measures from the peripheral auditory system in children who will not or cannot provide audiometric information.

Before DPOAEs can be used effectively as a clinical tool in the pediatric population, their basic characteristics must be defined. To date, there are few published studies about the characteristics of DPOAEs measured from significant numbers of infants and children. Based on the studies that exist, it is not clear if there are developmental changes in DPOAEs. Because there is strong evidence that COAEs decrease in level and change in frequency content with development (Norton and Widen, 1990; Prieve, 1992; Prieve *et al.*, 1997), it is possible that there are changes in DPOAEs with development as well.

The majority of studies suggest that DPOAEs from neonates (defined as infants aged less than one month) are different than those measured from adults; however, the details of how they differ vary among the studies. Lafreniere *et al.* (1991) and Smurzynski *et al.* (1993) measured DPOAEs using fixed-level primaries and plotted the DPOAE level as a function of the geometric mean frequency of the primaries. Both studies reported that the functions from neonates were ''qualitatively similar'' to those from adults; however, mean DPOAE levels from neonates were higher than those measured from adults for primary frequencies 3000 Hz and lower. For primary frequencies including and higher than 3400 Hz, DPOAE levels in neonates were lower than they were in adults. No statistical analyses were reported for these data to test for significant differences. Lasky *et al.* (1992) measured DPOAEs at 14 $f2$ frequencies ranging from 537 to 10 009 Hz using equal, high-level primaries in ten adults and ten neonates. Based on *t* tests, neonates had significantly higher DPOAE levels than adults at $f2$ frequencies of 2393 and 2783 Hz. No significant differences were found for the other 12 $f2$ frequencies.

Three studies indicated that DPOAE levels were higher than those from adults over a broad range of frequencies. Abdala and Sininger (1996) noted that DPOAE levels from neonates were 3–5 dB greater than those from adults at $f2$ frequencies of 1500, 3000, and 6000 Hz, but performed no statistical analysis to determine if they were significantly higher. Brown *et al.* (1994) reported that DPOAE levels in neonates were approximately 10 dB higher than they were in

a group of adults for $f2$ frequencies ranging from 1000 to 8000 Hz, but only the difference at the $f2$ frequency of 1000 Hz was significant. Bergman *et al.* (1995) also reported that DPOAE levels measured from graduates of the neonatal intensive care unit (NICU) and term neonates were higher than those from a group of older children and adults across a broad frequency span, but did not indicate whether these differences were statistically significant. In contrast to these studies, Popelka *et al.* (1995) reported that DPOAE input/output (I/O) functions from 19 healthy NICU infants were ''qualitatively and quantitatively similar'' to those from adult subjects, although direct comparisons of the two groups were not provided in their manuscript. Overall, the data from most studies suggest that neonates have higher DPOAE levels than adults; however, the details of the age-related differences are not consistent across studies. Furthermore, statistical analyses were not performed in most of the studies.

There have been few reports on DPOAEs from children. Those that exist suggest that there may be age-related differences in DPOAE levels between children and adults. Spektor *et al.* (1991) reported that DPOAE levels evoked by $f2$ frequencies ranging from 500 to 1300 Hz and at 6000 Hz were higher from a group of five children aged 4–10 yr than in adults. Lonsbury-Martin *et al.* (1994) also reported that children aged 2–14 yr had slightly higher DPOAE levels than adults for $f2$ frequencies between 4000 and 8000 Hz.

Although results from these studies of DPOAEs in neonates and children suggest that there are changes in level change with age, the question has not been thoroughly researched. Most studies employed small numbers of subjects and no statistical analyses were performed to determine if there were significant differences among age groups. Because the data suggest that there may be developmental differences in DPOAEs, together with strong evidence that COAEs (Norton and Widen, 1990; Prieve *et al.*, 1997) and spontaneous otoacoustic emissions (Burns *et al.*, 1994) change with development, there is a need for systematic study of DPOAEs from children of various ages.

In order to adequately study changes in DPOAEs with age, spontaneous otoacoustic emissions (SOAEs) must be monitored. It is has been reported that individuals with SOAEs have high COAE levels, low COAE ''thresholds,'' and I/O functions that saturate at low stimulus levels (Prieve and Falter, 1995; Probst *et al.*, 1986; Zwicker, 1983). DPOAEs are also influenced by SOAEs that are similar in frequency (Furst *et al.*, 1988; Gaskill and Brown, 1990; He and Schmiedt, 1993; Kemp, 1979; Long *et al.*, 1993; Moulin *et al.*, 1992; Wier *et al.*, 1988). However, it is not known whether the mean DPOAE level for a group of subjects having SOAEs is higher than that for a group of subjects who do not have SOAEs. If groups of individuals having SOAEs have higher DPOAE levels than those not having SOAEs, then SOAE status must be taken into account before determining whether DPOAEs are different among age groups.

The method of DPOAE data collection must also be considered when studying age-related changes. In many of the previous studies, DPOAEs were measured as a function of frequency with fixed-level primaries (Bergman *et al.*, 1995; Brown *et al.*, 1994; Lafreniere *et al.*, 1991; Lasky *et al.*, 1992; Smurzynski *et al.*, 1993; Smurzynski, 1994). However, evidence from gerbils indicates that DPOAE I/O functions show different patterns with development (Norton *et al.*, 1991). Variation in DPOAE I/O functions with development may imply that changes in the underlying mechanism for DPOAE generation is also changing. Although the degree to which the development of the auditory system in gerbils reflects that of humans is uncertain, the data from Norton *et al.* (1991) suggest that developmental changes in DPOAEs may be best explored through measuring I/O functions.

The purpose of the following study was to systematically study DPOAE I/O functions in children of various ages. The first goal was to determine if DPOAE levels change with childhood age; and if so, to define the changes. A second goal was to ensure that differences in the prevalence of SOAEs among the groups did not confound any possible developmental effects.

I. METHODS

DPOAEs were measured from one ear of 196 subjects aged four weeks to 29 yr. Data from subjects were placed into one of seven groups based on age: Less than 1 yr ($n=28$, $\overline{x}=0.51$, s.d.$=0.23$,) 1–3 yr ($n=28$, $\overline{x}=2.72$, s.d.$=0.63$,) 4–5 yr ($n=24$, $\overline{x}=4.59$, s.d.$=0.56$,) 6–8 yr ($n=30$, $\overline{x}=6.87$, s.d.$=0.81$,) 9–11 yr ($n=23$, $\overline{x}=9.72$, s.d.$=0.79$,) 12–17 yr ($n=28$, $\overline{x}=13.61$, s.d.$=1.74$,) and 18–29 yr ($n=35$, $\overline{x}=22.8$, s.d.$=1.85$.) Age groups were the same as those used by Prieve *et al.* (1997) so that any differences in DPOAE level among age groups could be compared to those observed for COAEs. Descriptions of audiometric and tympanometric testing procedures can be found in Prieve *et al.* (1997). Seventy-two percent of the subjects who participated in the study by Prieve *et al.* (1997) also participated in the present study.

The ILO88XP system in conjunction with a personal computer was used to measure DPOAEs. The primary tones were generated via separate 14-bit digital-to-analog converters with sampling rates of 25 kHz. The primary tones were transduced by separate receivers (Knowles BP 1712), which were housed along with a microphone (Knowles EA 1843) in the same probe. The ''adult'' probe was used to measure DPOAEs in all subjects. Cross-talk of the primaries measured by the probe microphone is reported to be 60 dB for frequencies less than 8000 Hz (Siegel, 1995).

DPOAE I/O functions were measured with $f2$ frequencies of 1000, 1500, 2000, 3000, 4000, 5000, and 6000 Hz. The $f2{:}f1$ ratio was 1.22 and $f1$ level was presented 15 dB higher than $f2$. The $f2$ levels ranged from 40 to 60 dB SPL for $f2$ frequencies at 1000, 1500, 2000, 3000, and 4000 Hz and from 30 to 60 dB SPL for $f2$ frequencies at 5000 and 6000 Hz.

Primary stimuli were calibrated in the ear canal and the calibration was completed for each $f2$ frequency prior to measurement of its growth function. The calibration procedure was divided into two steps. The first step provided feedback about the quality of the probe fit in the ear canal. Three stimuli were presented. The first was a 250-Hz sine wave, whose amplitude was measured by the probe microphone.

The software calculated the effective ear canal volume based on the amplitude together with data stored in a probe calibration file. The second and third stimuli were broadband click stimuli, which were presented to each output transducer. A real-time fast Fourier transformation (FFT) of each stimulus in the ear canal was displayed so that the operator could achieve the flattest response across the frequency range. The system compensated for level differences across frequency that were within $+6$ to -9 dB from that at 2000 Hz. The system did not compensate for large dips or peaks in the spectrum because they are likely caused by standing waves and correction for level would result in large errors. In the second step of the calibration procedure, the two primaries were set to the requested levels. The operator also adjusted the sensitivity of system to reject incoming data containing excessive noise. The incoming signal was checked for excessive noise after an FFT had been performed on it and the stimulus energy had been removed. The primary stimuli were presented initially at the highest levels and decreased in steps of 5–10 dB to the lowest level.

The sound in the ear canal was sampled via a 14-bit, analog-to-digital converter at a rate of 25 kHz. The data record was 80 ms in length and consisted of 2024 points. Incoming data were synchronously averaged in the time domain. Each data point was based on approximately 240 averages (range was from 200 to 280 averages), collected in subaverages of 16 samples. The level at the $2f1-f2$ frequency was calculated from an FFT having a frequency resolution of 12.5 Hz, which was performed on the averaged trace and updated after each subaverage. The average and standard deviation of the ten adjacent spectral lines (half above and half below) to the $2f1-f2$ frequency were calculated to estimate the noise. The noise was displayed in $+1$ and $+2$ standard deviations from its mean. A DPOAE was considered present and added to the data base if the primary stimulus levels were within ± 3 dB of the requested stimulus level and the DPOAE was at least 2 dB above two standard deviations of the noise.

Similar to Prieve *et al.* (1997), test procedures were designed to accommodate infants and young children as well as the adults. DPOAEs from infants and young children were measured in 10-dB steps at 2000, 4000, and 6000 Hz. If the child was still cooperative, 1000, 1500, 3000, and 5000 Hz were also measured. DPOAEs were collected in 5 dB steps at all frequencies in subjects aged 6 yr and older. Because many of the subjects also participated in the study reported by Prieve *et al.* (1997), SOAEs had already been measured. If the child had not participated in the study mentioned above, SOAEs were measured by synchronizing them to a click using the procedures described in Prieve *et al.* (1997). Descriptions of subject awareness during testing can be found in Prieve *et al.* (1997).

II. RESULTS

Because other researchers have reported various patterns of DPOAE I/O functions among adults (He and Schmiedt, 1993; Nelson and Kimberley, 1992; Popelka *et al.*, 1993; Stover and Norton, 1993) and neonates (Popelka *et al.*, 1995), individual I/O functions were assessed in all subjects.

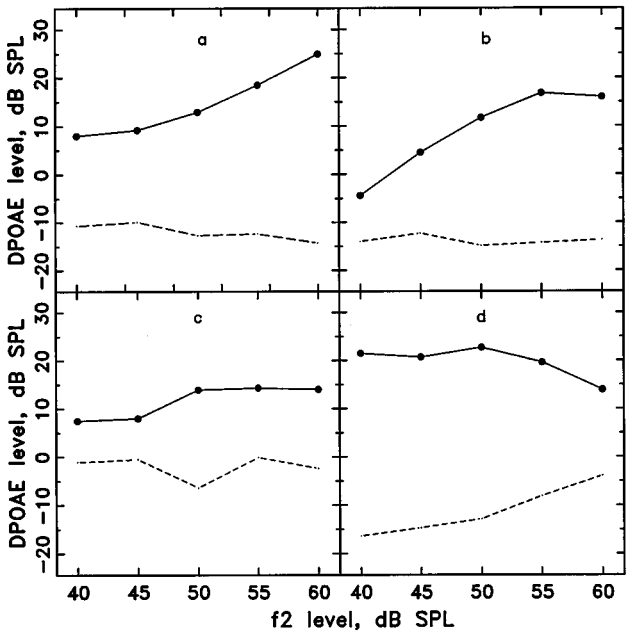


FIG. 1. Examples of DPOAE I/O functions in four individuals. Seventy-eight percent of the functions were increasing or saturating functions similar to those illustrated in panels (a) and (b). Only 3% were flat, as illustrated in panel (c). Nineteen percent were nonmonotonic functions that exhibited lower DPOAE levels at higher stimulus levels (illustrated in panel d) or exhibited a ''dip'' or flat section of the function at midstimulus levels. The dashed lines represents two standard deviations higher than the mean noise (see methods for description).

The I/O functions were placed into three broad categories: increasing (monotonically increasing or increasing with saturation at high stimulus levels), flat (all data points were within ± 3 dB of each other), or nonmonotonic (flat portions present for low- or midstimulus levels, a ''dip'' present for midlevel stimuli, or the function was flat and decreased at high stimulus levels). Only three categories were used in the classification because detailed I/O functions were not measured. In older subjects, I/O functions spanned only 20–30 dB and in those aged 5 yr and less, usually only three data points were collected per function. Figure 1 illustrates typical I/O functions from individual subjects. Seventy-eight percent of the individual I/O functions were considered ''increasing'' and examples are shown in panels 1(a) and 1(b). Three percent of the I/O functions were flat, as shown in panel 1(c). This type of function was noted most frequently for $f2$ frequencies of 1000 and 1500 Hz. Nineteen percent of the functions were judged to be nonmonotonic [panel 1(d)]. The proportion of occurrence of the four patterns was similar among age groups.

Figures 2–4 illustrate the mean DPOAE level as a function of $f2$ level. The parameter is subject age group, each represented by a different symbol. Figure 2 illustrates mean DPOAE I/O functions for $f2$ frequencies of 1000 and 1500 Hz. Figure 3 illustrates mean DPOAE I/O functions for $f2$ frequencies of 2000 and 3000 Hz and Fig. 4 illustrates the mean DPOAE I/O function for 4000, 5000, and 6000 Hz. The long-dashed line represents the average of two standard deviations higher than the mean individual noise for adults and the dotted line represents the corollary for children aged less than 1 yr. The open star symbol connected by the short-

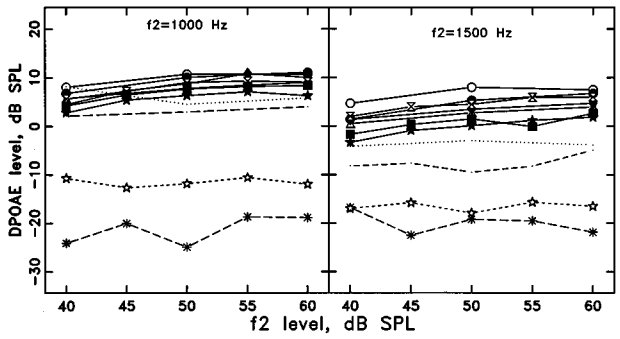


FIG. 2. Mean DPOAE I/O functions for $f2$ frequencies of 1000 and 1500 Hz. Subject age group is the parameter (<1 yr, open circle; 1–3 yr, closed circle; 4 to 5 yr, open triangle; 6–8 yr, closed diamond; 9–11 yr, open hourglass; 12–17 yr, closed square; adults, closed star). The dotted line represents the average of two standard deviations above the noise for children aged <1 yr and the long-dashed line represents the corollary for adults. The open stars connected by the short-dashed line represents the mean noise from ten runs of the experimental protocol in a 1-cc cavity. The asterisks connected by long-dashed lines represent the mean level at the $2f1-f2$ frequency from ten runs of the experimental protocol in a 1-cc cavity. Standard deviations range from 3.48 to 6.48 dB at 1000 Hz and from 4.35 to 7.77 dB at 1500 Hz.

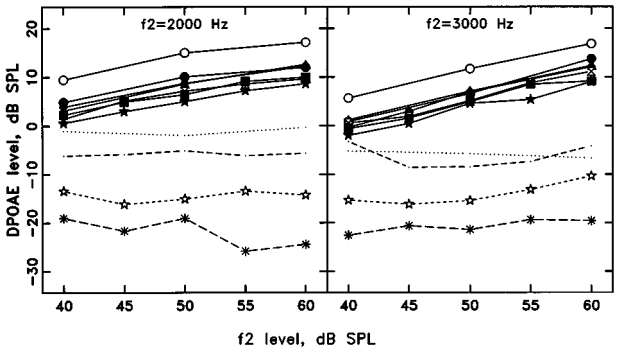


FIG. 3. Mean DPOAE I/O functions for $f2$ frequencies of 2000 and 3000 Hz. Details are given in the legend of Fig. 2. Standard deviations range from 3.55 to 7.4 dB at 2000 Hz and from 4.2 to 7.4 dB at 3000 Hz.

dashed line represents the average noise level two standard deviations above the average noise when the test was run in a 1-cc cavity (ten runs were performed). Finally, the asterisks connected by the dashed line represents the mean sound-pressure level at the calculated $2f1-f2$ frequency with the probe placed in a 1-cc cavity (ten runs).

For most $f2$ frequencies, DPOAE levels were different among age groups, with the children aged less than 1 yr having the highest DPOAE levels. To determine if there were significant differences among groups, a 2 (SOAE status)×7(group)×3 ($f2$ level of 40, 50, and 60 dB SPL) mixed-model ANOVA with repeated measures on the last factor were performed for each $f2$ frequency.

There were significant differences in DPOAE levels across group for every frequency except 1000 Hz. To determine which groups were significantly different than others, *post-hoc* Tukey tests were performed ($\alpha=0.05$ familywise). Table I lists the significant main effects for each ANOVA and the significant differences between groups based on the Tukey tests. Children aged less than 1 yr had higher mean DPOAE levels than all other groups at 2000 and 3000 Hz. They also had higher mean DPOAE levels than teens aged 12–17 yr and adults at 1500, 4000, and 5000 Hz and higher levels than adults at 6000 Hz. Children aged 1–3 yr had higher DPOAE levels than adults at 3000, 5000, and 6000 Hz. At 6000 Hz children aged 4 to 5 and 6–8 yr also had higher DPOAE levels than adults.

The shapes of the mean functions change with frequency. At 1000 and 1500 Hz the functions are relatively flat. The functions have an increasing slope at 2000 and 3000 Hz. For 4000–6000 Hz, the mean functions have linear portions but begin to saturate for higher stimulus levels. The main effect of the stimulus level was statistically significant at every frequency ($p<0.00005$ at each frequency). Because the trend of increasing DPOAE level with increasing stimulus level is not surprising, follow-up tests were not performed on this main factor. There were no group×level interactions (p values ranged from 0.09 to 0.80), indicating that differences between groups were similar regardless of stimulus intensity.

For all frequencies, DPOAE levels for individuals hav-

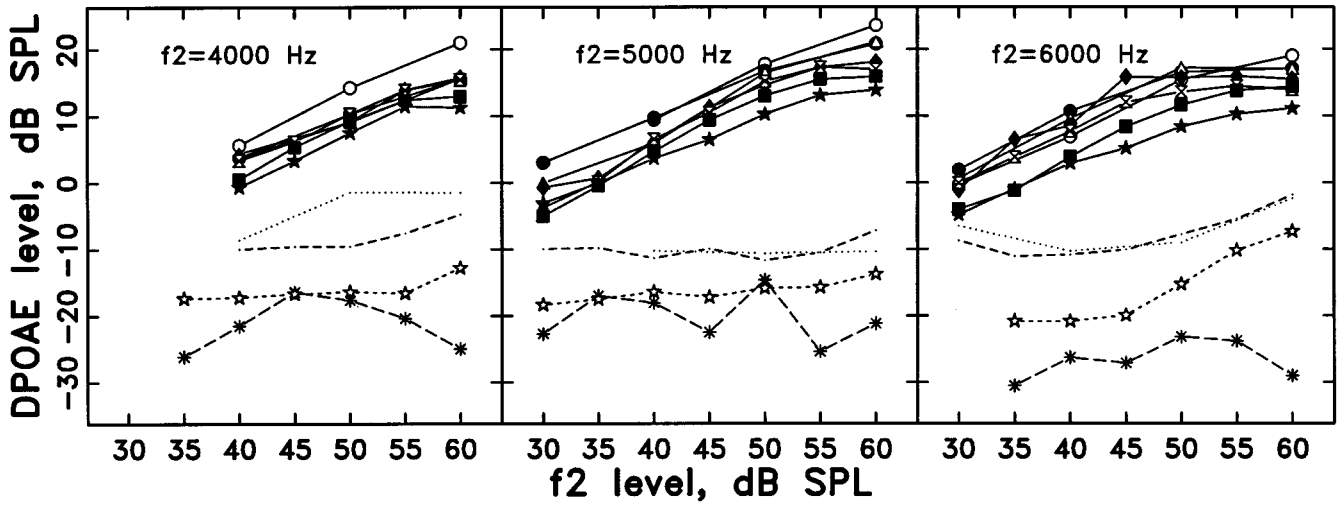


FIG. 4. Mean DPOAE I/O functions at the $f2$ frequency of 4000, 5000, and 6000 Hz. Details are given in the legend of Fig. 2. Standard deviations range from 4.34 to 7.6 dB at 4000 Hz, from 3.0 to 8.0 dB at 5000 Hz and from 3.8 to 8.8 dB at 6000 Hz.

TABLE I. Significant main effects of the 7(group)×2 (SOAE status)×3 (stimulus level) mixed-model ANOVA, with repeated measures on the last factor. An ANOVA was performed for each $f2$ frequency. There were no significant interactions for any frequency. Significant differences in mean DPOAE levels among groups was performed *post-hoc* via a Tukey test ($\alpha = 0.05$ familywise) and are listed in the second column.

$f2$ frequency	Sig. main effects	Sig. follow-up on group differences
1000 Hz	SOAE,[a] Level[b]	none performed
1500 Hz	SOAE,[c] Group,[c] Level[b]	<1 yr>4 to 5 yr, 12–17 yr, adults
2000 Hz	SOAE,[b] Group,[b] Level[b]	<1 yr>1–17 yr, adults
3000 Hz	SOAE,[b] Group,[b] Level[b]	<1 yr>1–17 yr, adults 1–3 yr>adults
4000 Hz	SOAE,[b] Group,[b] Level[b]	<1 yr>12–17 yr, adults 9–11 yr>adults
5000 Hz	SOAE,[a] Group,[c] Level[b]	<1 yr>12–17 yr, adults 1–3 yr>adults
6000 Hz	SOAE,[c] Group,[b] Level[b]	<1 yr>adults 1–3 yr>2–17 yr, adults 4 to 5 yr>adults 6–8 yr>adults

[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.000\,05$.

ing SOAEs were significantly higher than for those not having SOAEs. The differences in mean DPOAE levels were approximately 4 dB for $f2$ frequencies of 2000, 3000, and 4000 Hz, 3 dB at 1500 Hz and approximately 2 dB for $f2$ frequencies of 1000, 5000, and 6000 Hz. There were no significant SOAE×level, SOAE×group, or SOAE ×group×level interactions (p values ranged from 0.11 to 0.94). Figure 5 illustrates the comparison of mean I/O functions between subjects with measurable SOAEs and subjects without measurable SOAEs for all age groups for an $f2$ frequency of 1000 Hz. Figure 6 provides a second example of this comparison for an $f2$ frequency of 3000 Hz.

III. DISCUSSION

Similar to reports from other researchers (Lonsbury-Martin *et al.*, 1990; Nelson and Kimberley, 1992; Popelka *et al.*, 1993, 1995; Stover and Norton, 1993), individual

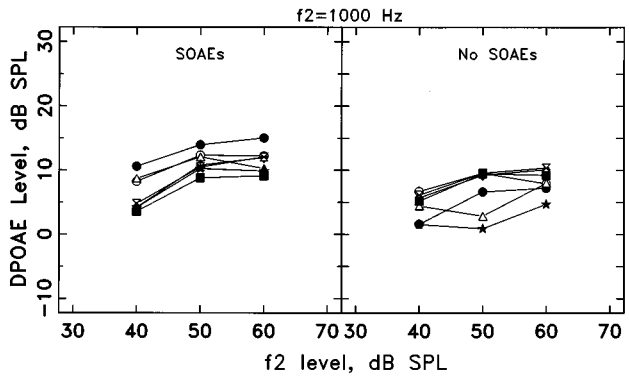


FIG. 5. Mean DPOAE I/O functions at the $f2$ frequency of 1000 Hz for individuals having SOAEs (left panel) and without measurable SOAEs (right panel). There were no significant differences in DPOAE levels among age groups, but mean DPOAE level was significantly higher (2 dB) for individuals having SOAEs than for those not having measurable SOAEs.

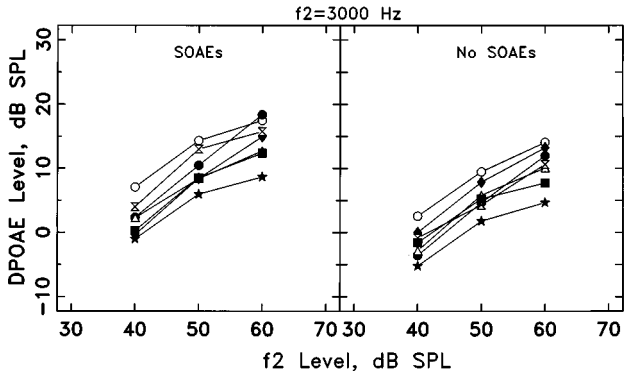


FIG. 6. Mean DPOAE I/O functions at the $f2$ frequency of 3000 Hz. The mean DPOAE level was significantly higher (3.77 dB) for individuals having SOAEs than for those not having SOAEs. Significant differences among groups are listed in Table I.

DPOAE I/O functions exhibited various patterns of growth. Therefore the shapes of mean I/O functions do not reflect every individual I/O function. Thus some researchers have opted to not average I/O functions and instead, classify them based on their shape (Nelson and Kimberley, 1992; Popelka *et al.*, 1993, 1995). In the current study, it was only practical to classify shapes into very broad categories because DPOAEs were collected over a narrow range of stimulus levels and because a large step size, 10 dB, was used in young subjects. Although detailed shapes were not classified, 78% of the individual I/O functions were judged to be monotonically increasing or monotonically increasing with saturation. Although it is still uncertain why individuals exhibit various types of DPOAE functions (He and Schmiedt, 1993; Long *et al.*, 1993), the different patterns were seen in the same proportions in all age groups. Because the age groups exhibited all patterns in the same proportions, there is no reason to believe that the various individual I/O functions averaged differently across groups.

The main finding of the present study is that there were differences in mean DPOAE levels among age groups. The differences were frequency dependent, but independent of stimulus level. Infants had higher mean DPOAE levels for $f2$ frequencies ranging from 1500 to 5000 Hz than teens aged 12–17 yr and adults. At 2000 and 3000 Hz, infant mean DPOAE levels were higher than those from all other age groups. Children aged 1–3 yr had higher mean DPOAE levels than young adults at 3000, 5000, and 6000 Hz. At 6000 Hz, children aged less than 8 yr had higher mean DPOAE levels than adults. Children aged 6–8 yr had higher DPOAE levels than adults at 3000 Hz and those aged 9–11 yr had higher DPOAE levels than did adults at 4000 Hz.

There are few published data to which the results of the current study can be directly compared because of two factors: differences in subject population and stimulus parameters. Most of the published data from infants have been collected on neonates (Abdala and Sininger, 1996; Brown *et al.*, 1994; Lafreniere *et al.*, 1991; Lasky *et al.*, 1992; Smurzynski *et al.*, 1993) or NICU graduates (Bergman *et al.*, 1995; Popelka *et al.*, 1995; Smurzynski, 1994) as opposed to older infants, as was the case in the current study. In some of these studies, DPOAE levels from neonates were reported to

be higher than those from adults for $f2$ frequencies ranging from 1600 to 3000 Hz (Lafreneire *et al.*, 1991; Lasky *et al.*, 1992; Smurzynski *et al.*, 1993) whereas others reported that neonates had higher DPOAE levels than adults for $f2$ frequencies ranging from 1000 to 8000 Hz (Abdala and Sininger, 1996; Brown *et al.*, 1994; Bergman *et al.*, 1995). DPOAE levels for infants in the current study were higher than those for adults for $f2$ frequencies from 1500 to 6000 Hz. In addition, they were also higher than those in older children over much of this frequency range.

Stimulus levels used to evoke DPOAEs in the current study were different from most of the other published studies, in which equal-level primaries were used. With respect to stimulus parameters, the current study is most similar to those conducted by Brown *et al.* (1994) and Bergman *et al.* (1995). DPOAE levels in these studies were evoked by $f2$ primary levels of 40 dB SPL and 50 dB SPL, respectively, with $f1$ primary levels being 15 dB higher than $f2$ primary levels. To better compare results from the current study to those from these previously published reports, the data have been replotted for the DPOAE level as a function of $f2$ frequency in Fig. 7. Mean DPOAE levels evoked by $f2$ stimulus levels of 40 dB SPL for infants aged less than 1 yr and adults are shown in the top panel and those evoked by $f2$ stimulus levels of 50 dB SPL are shown in the bottom panel. The data from Brown *et al.* (1994), collected from term neonates and adults, are illustrated in the top panel and those from Bergman *et al.* (1995), collected from a group of NICU graduates (aged 35–45 conceptional weeks) and a group of ''older children'' and adults, are shown in the bottom panel. Mean adult DPOAE levels from adults reported by Brown *et al.* (1994) are nearly identical to those from the present study for $f2$ frequencies from 1000 to 6000 Hz and mean adult DPOAE levels from Bergman *et al.* (1995) are within one standard deviation of those from the current study for $f2$ frequencies from 1000 to 4000 Hz. Mean adult DPOAE levels for $f2$ frequencies at 5000 and 6000 Hz are higher than those reported by Bergman *et al.* (1995). Mean DPOAE levels evoked by $f2$ levels of 40 dB SPL from infants are similar to those from neonates from 2000 to 6000 Hz (top panel). Mean DPOAE levels evoked by $f2$ levels of 50 dB SPL from infants are similar to those from NICU graduates for $f2$ frequencies at 1000 and 1500 Hz, but higher for $f2$ frequencies from 3000 to 6000 Hz. The differences in DPOAE levels among the three studies could be attributed to at least two factors.

The first factor that could account for some of the differences in DPOAE levels is that different instrumentation was used in all three studies. The best way to address whether instrumentation accounts for differences in DPOAE level is to compare data from the adult groups. Mean adult DPOAE levels from the current study and Brown *et al.* (1994) are nearly identical. The shapes of the functions for adult groups from the current study and Bergman *et al.* (1995) are similar and the mean DPOAE levels are similar for $f2$ frequencies from 1000 to 4000 Hz. However, the mean DPOAE levels for $f2$ frequencies at 5000 and 6000 Hz from the current study are higher than those reported by Bergman *et al.* (1995). These differences could be due to

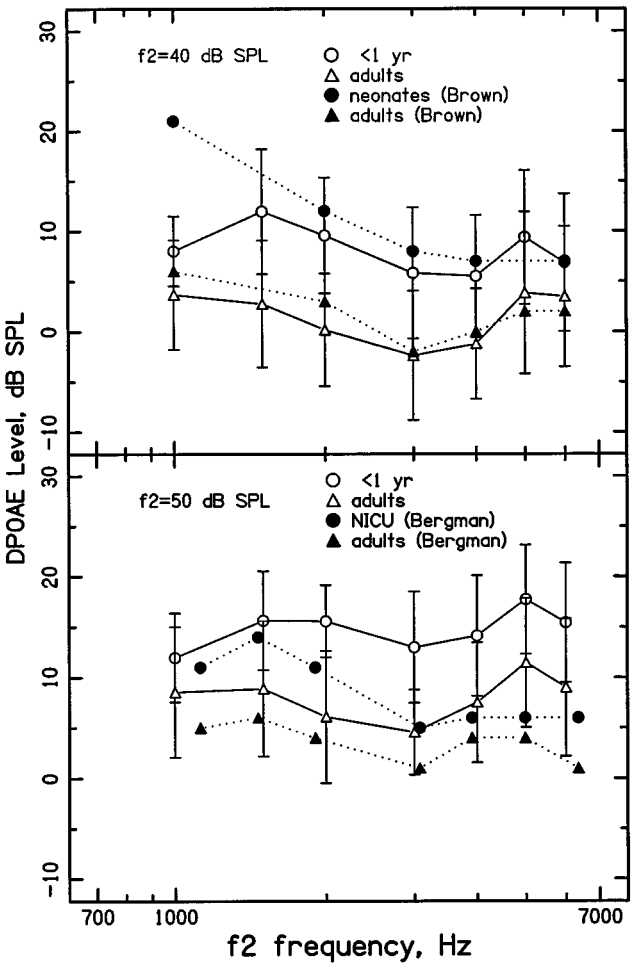


FIG. 7. Mean DPOAE levels evoked by $f2$ levels of 40 dB SPL (top panel) and 50 dB SPL (bottom panel) plotted as a function of $f2$ frequency for infants (open circles) and adults (open triangles) from the current study. The $f1$ levels were 15 dB higher then the $f2$ levels. Error bars represent ± 1 standard deviation. Mean DPOAE levels evoked by similar primary stimuli from term neonates (closed circles) and adults (closed triangles) from Brown *et al.* (1994) are illustrated in the top panel. Those from NICU graduates (closed circles) and a group of ''older'' children and adults (closed triangles) from Bergman *et al.* (1995) are shown in the bottom panel.

equipment differences, such as probe impedance (e.g., Lafreniere *et al.*, 1991), or to calibration error at high frequencies. Data for the Etymotic 10B microphone/earphone system used by Bergman *et al.* (1995) indicate that there can be major errors in the sound-pressure level at the eardrum for frequencies at and higher than 4000 Hz when it is calibrated at the probe (Siegel, 1994; Siegel and Hirohata, 1994). No such data exist for the ILO88 XP probe system; however, the manual states that calibration at 4000 Hz and above can contain inaccuracies due to standing waves. Because possible calibration inaccuracies will be apparent for high frequencies, it is possible that the differences in mean DPOAE level between the adult groups from the two studies are due to differences in calibration errors.

The second factor that could contribute to the differences in DPOAE level among the infant groups is that different populations were studied. Brown *et al.* (1994) measured DPOAEs from term neonates and their mean levels are similar to those from infants, except for the $f2$ frequency at

1000 Hz. Because calibration errors are not typically a problem at 1000 Hz, the differences could be attributable to the populations being tested. DPOAE levels from NICU graduates (Bergman *et al.*, 1995) are considerably lower than those from infants from 2000 to 6000 Hz. Based on data reported by Smurzynski (1994), it is possible that infants aged 45–52 weeks have higher mean DPOAE levels than NICU graduates aged 35–45 weeks (conceptional age). Smurzynski (1994) longitudinally measured DPOAEs from infants aged 33–41 weeks who were patients in the NICU. He found that DPOAE levels evoked by stimuli having frequencies from 2.8 to 4 kHz increased in 77%–85% of ears and those evoked by stimuli at 1.4 and 5.6 kHz increased in 50% of ears with conceptional age. It is over these same frequency ranges where mean DPOAE levels from infants (current study) are higher than those from NICU graduates (Bergman *et al.*, 1995).

Two studies exist in which DPOAEs were measured in children, and results from the current study are similar to one of them. Lonsbury-Martin *et al.* (1994) found that 43 ears from a group of normally hearing children aged 2–14 yr had slightly higher DPOAE levels from 4000 to 8000 Hz than those from a group of normally hearing adults. They did not provide statistical analysis to determine if the means were significantly different. Results from the current study indicated that children aged 1–3 yr had significantly higher DPOAE levels than adults at 3000, 5000, and 6000 Hz. Children aged 4 to 5 and 6–8 yr also had higher DPOAE levels than adults at 6000 Hz. The results from the current study differ from those reported by Spektor *et al.* (1991), who found that children aged 4–10 yr had higher DPOAE levels than adults from 700 to 1400 Hz. The differences may be attributable, at least in part, to the fact that Spektor and colleagues used equal-level primaries to evoke DPOAEs and that only five normally hearing children were tested.

Another outcome of the present study was that mean DPOAE levels were 2–4 dB higher for groups of individuals having SOAEs than for groups not having SOAEs. Although these mean differences were small, they were statistically significant for every $f2$ frequency. Previous research has shown that in individual subjects, the DPOAE level was typically higher when its frequency was close to that of an SOAE (Furst *et al.*, 1988; Gaskill and Brown, 1990; He and Schmiedt, 1993; Kemp, 1979; Long *et al.*, 1993; Moulin *et al.*, 1992; Wier *et al.*, 1988). For the ANOVAs run in the current study, only the condition of whether an individual did or did not have an SOAE was recorded, without any regard to the SOAE frequencies or levels. Yet, for every $f2$ frequency, mean DPOAE levels were higher for groups of individuals having SOAEs than for those who did not. Inspection of individual SOAE data indicated that there were 409 SOAEs measured across subjects. Fifty-nine of the SOAEs were within ± 80 Hz of a $2f1-f2$ DPOAE frequency. It is uncertain if these few SOAEs would alter individual DPOAE levels enough to increase the mean DPOAE level.

Another possible way in which an SOAE could interact with a DPOAE, is if its frequency was close to one of the primary frequencies. Although SOAEs are suppressed by pure tones of surrounding frequencies (Bargones and Burns, 1988; Long *et al.*, 1991; Zizz and Glattke, 1988), a SOAE will synchronize to pure tones that are very close in frequency. When the SOAE is synchronized, it tends to increase the level of the pure tone as measured in the ear canal. The frequency range over which SOAEs synchronize to an external tone is dependent on the level of the tone, but is generally on the order of 30 Hz (Long *et al.*, 1991; Zwicker and Schloth, 1984). It is possible that SOAEs close in frequency to one of the primaries influenced the action of the primary creating the DPOAE, resulting in an increased DPOAE level. In the present study, a total of 60 SOAEs were within ± 30 Hz of a primary frequency. Therefore 29% of SOAEs were close to the DPOAE frequency, the $f1$ frequency, or the $f2$ frequency. Again, it is uncertain if individual DPOAE levels altered by the SOAEs would account for an increased mean DPOAE level.

Interestingly, few individuals had SOAEs having frequencies of 4000 Hz and higher, yet mean DPOAE levels were significantly higher for $f2$ frequencies of 4000, 5000, and 6000 Hz for groups of individuals having SOAEs than for those not having SOAEs. This suggests that perhaps SOAEs are indicative of a ''more active'' cochlea in general and it is that condition that influences the level of the DPOAE as well as the SOAE itself. Similar thinking has been proposed by Kummer *et al.* (1995) and Wier *et al.* (1988). Wier *et al.* (1988) found that DPOAEs close in frequency to SOAEs were high in level, even when aspirin consumption had reduced the SOAE so that it was unmeasurable. They concluded that the features of the SOAE generator mechanism remain active even when an SOAE is not measurable. Although they limited their discussion to generators at specific places in the cochlea, perhaps it is possible that the presence of ''active places'' impact DPOAE generation in general.

There were significant differences in mean DPOAE levels among age groups regardless of whether $f2$ level was 40, 50, or 60 dB SPL. This was shown by the lack of significant group $\times$ level interactions. Although the slopes of individual and mean I/O functions were not calculated, the finding that there were no significant group $\times$ level interactions suggests that the slope of mean I/O functions was not different among groups. These results are different than I/O functions for developing gerbils, in which slope and shape of I/O functions approximated those of adults only with development (Norton *et al.*, 1991).

Differences in mean DPOAE levels among age groups were similar to those seen for COAE levels among age groups. A comparison between significant differences in the DPOAE level and COAE level across age groups is given in Table II (COAE data from Prieve *et al.*, 1997). For both types of EOAEs, there were no significant differences among groups at 1000 Hz. At 1500 Hz infants had higher EOAE levels than subjects aged 12–17 yr and adults. Infants had higher levels than all other subjects for both types of EOAEs at 2000 and 3000 Hz (3170 Hz for COAEs). At 4000 and 5000 Hz (5040 Hz for COAEs), infants had higher levels than those aged 12–17 yr and adults. Children aged 1–3 yr had higher EOAE levels than adults at 3000 and 5000 Hz.

TABLE II. Comparison of significant differences between groups for COAEs parsed into 1/3-octave bands and DPOAEs at a corresponding frequency. COAEs were evoked by 80 dB pSPL clicks and measured using a subtraction paradigm. Data for the COAEs are detailed by Prieve *et al.* (1997).

Frequency	COAE	DPOAE
1000	none	none
1500	<1 yr>4–17 yr, adults	<1 yr>4 to 5 yr,
(1590 for COAEs)		12–17 yr, adults
2000	<1 yr>1–17 yr, adults	<1 yr>1–17 yr, adults
3000	<1 yr>1–17 yr, adults	<1 yr>1–17 yr, adults
(3170 for COAEs)	1–3 yr>12–17 yr, adults	1–3 yr>adults
4000	<1 yr>4–17 yr, adults	<1 yr>12–17 yr, adults
	1–3 yr>12–17 yr, adults	9–11 yr>adults
5000	<1 yr>4–17 yr, adults	<1 yr>12–17 yr, adults
(5040 for COAEs)	1–3 yr>4–17 yr	1–3 yr>adults

There were no COAE data for comparison at 6000 Hz because it was not possible to collect data in that frequency range. At 6000 Hz, children aged less than 8 yr had higher DPOAE levels than adults. It is unsure whether DPOAE levels at 6000 Hz are due to calibration differences, actual differences in DPOAE levels, or a combination of factors. To summarize, there is close similarity in maturational change of DPOAEs and COAEs from 1000 to 5000 Hz.

It is not surprising that both types of EOAE levels decrease with maturation in a similar manner as both arise from cochlear processes. One explanation for change with age is that generators or processes of generation for both types of EOAEs are in place early in life, and that the changes are due to maturation of the middle ear and ear canal. Once effects of the middle ear and ear canal have been accounted for, it can be investigated if any changes in EOAEs with age could be due to cochlear maturation.

Because mean DPOAE levels in infants are approximately 10 dB higher than they are in adults for some frequencies, it may be necessary to use separate norms for infants and young children. However, further research including DPOAEs from hearing-impaired children is needed to determine if clinical criteria for infants and children will be different from those used for adults.

ACKNOWLEDGMENTS

Thanks are given to Lara Allesandrelli, Patti Baran Kopylczak, Heather Grossman, Jennifer Horn, Holly Huta, Ann Jackson, Tara McKinney, Nadia Michail, Margaret Ramstad, and Deb Rotman Dauber for their assistance with data collection and analysis. Dr. Karen Doherty, Dr. Ted Glattke, and an anonymous reviewer provided helpful comments on an earlier version of this manuscript. This work was funded by the Deafness Research Foundation, the American Speech-Language-Hearing Association Foundation and by NIDCD Grant No. 5 R29 DC02028.

Abdala, C., and Sininger, Y. S. (**1996**). ''The development of cochlear frequency resolution in the human auditory system,'' Ear Hear. **17**, 374–385.

Bargones, J. Y., and Burns, E. M. (**1988**). ''Suppression tuning curves for spontaneous otoacoustic emissions in infants and adults,'' J. Acoust. Soc. Am. **83**, 1809–1816.

Bergman, B. M., Gorga, M. P., Neely, S. T., Kaminski, J. R., Beauchaine, K. L., and Peters, J. (**1995**). ''Preliminary descriptions of transient-evoked and distortion-product otoacoustic emissions from graduates of an intensive care nursery,'' J. Am. Acad. Audiol. **6**, 150–162.

Brown, A. M., Sheppard, S. L., and Russell, P. T. (**1994**). ''Acoustic distortion products (ADP) from the ears of term infants and young adults using low stimulus levels,'' Br. J. Audiol. **28**, 273–280.

Burns, E. M., Campbell, S. L., and Arehart, K. H. (**1994**). ''Longitudinal measurements of spontaneous otoacoustic emissions in infants,'' J. Acoust. Soc. Am. **95**, 385–394.

Furst, M., Rabinowitz, W. M., and Zurek, P. M. (**1988**). ''Ear canal acoustic distortion at $2f1-f2$ from human ears: Relation to other emissions and perceived combination tones,'' J. Acoust. Soc. Am. **84**, 215–221.

Gaskill, S. A., and Brown, A. M. (**1990**). ''The behavior of the acoustic distortion product, $2f1-f2$, from the human ear and its relation to auditory sensitivity,'' J. Acoust. Soc. Am. **88**, 821–839.

He, N.-J., and Schmiedt, R. A. (**1993**). ''Fine structure of the $2f1-f2$ acoustic distortion product: Changes with primary level,'' J. Acoust. Soc. Am. **94**, 2659–2669.

Kemp, D. T. (**1979**). ''Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea,'' Arch. Otorhinolaryngol. **224**, 37–45.

Kummer, P., Janssen, T., and Arnold, W. (**1995**). ''Suppression tuning characteristics of the $2f1-f2$ distortion-product otoacoustic emission in humans,'' J. Acoust. Soc. Am. **98**, 197–210.

Lafreniere, D., Jung, J. D., Smurzynski, J., Leonard, G., Kim, D. O., and Sasek, J. (**1991**). ''Distortion-product and click-evoked otoacoustic emissions in healthy newborns,'' Arch. Otolaryngol. Head Neck Surgery **117**, 1382–1389.

Lasky, R., Perlman, J., and Hecox, K. (**1992**). ''Distortion-product otoacoustic emissions in human newborns and adults,'' Ear Hear. **13**, 430–441.

Long, G. R., Sun, C., and Talmadge, C. L. (**1993**). ''Interactions between spontaneous emissions and external tones: Suppression, frequency shifting and distortion product generation,'' in *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 40–46.

Long, G. R., Tubis, A., and Jones, K. L. (**1991**). ''Modeling synchronization and suppression of spontaneous otoacoustic emissions using Van der Pol oscillators: Effects of aspirin administration,'' J. Acoust. Soc. Am. **89**, 1201–1221.

Lonsbury-Martin, B. L., Harris, F. P., Stagner, B. B., Hawkins, M. D., and Martin, G. K. (**1990**). ''Distortion product emissions in humans. I. Basic properties in normally hearing subjects,'' Ann. Otol. Rhinol. Laryngol. Suppl. **147**, 3–14.

Lonsbury-Martin, B. L., Martin, G. K., McCoy, M. J., and Whitehead, M. L. (**1994**). ''Otoacoustic emissions testing in young children: Middle-ear influences,'' Am. J. Otology **15** (Suppl. 1), 13–20.

Moulin, A., Collet, L., and Morgon, A. (**1992**). ''Influence of spontaneous otoacoustic emissions (SOAE) on acoustic distortion product input/output functions: Does the medial efferent system act differently in the vicinity of an SOAE?'' Acta Oto-Laryngol. **112**, 210–214.

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.

Norton, S. J., Bargones, J. Y., and Rubel, E. W. (**1991**). ''Development of otoacoustic emissions in gerbil: Evidence for micromechanical changes underlying development of the place code,'' Hearing Res. **51**, 73–92.

Norton, S. J., and Widen, J. E. (**1990**). ''Evoked otoacoustic emissions in normal-hearing infants and children: Emerging data and issues,'' Ear Hear. **11**, 121–127.

Popelka, G. R., Karzon, R. K., and Arjmand, E. M. (**1995**). ''Growth of the $2f1-f2$ distortion product emission for low-level stimuli in human neonates,'' Ear Hear. **16**, 159–165.

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,'' Hearing Res. **71**, 12–22.

Prieve, B. A. (**1992**). ''Otoacoustic emissions in infants and children: Basic characteristics and clinical application,'' Seminars Hear. **13**, 37–52.

Prieve, B. A., and Falter, S. R. (**1995**). ''COAEs and SSOAEs in adults with increased age,'' Ear Hear. **16**, 521–528.

Prieve, B. A., Fitzgerald, T. S., and Schulte, L. E. (**1997**). ''Basic characteristics of click-evoked otoacoustic emissions in infants and children,'' J. Acoust. Soc. Am. **102**, 2860–2870.

Probst, R., Lonsbury-Martin, B. L., Martin, G. K., and Coats, A. C. (**1986**). ''Otoacoustic emissions in ears with hearing loss,'' Am. J. Otolaryngol. **8**, 73–81.

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. (**1995**). ''Cross-talk in otoacoustic emission probes,'' Ear Hear. **16**, 150–158.

Siegel, J. H., and Hirohata, E. T. (**1994**). ''Sound calibration and distortion product otoacoustic emissions at high frequencies,'' Hearing Res. **80**, 146–152.

Smurzynski, J. (**1994**). ''Longitudinal measurements of distortion-product and click-evoked otoacoustic emissions of preterm infants: Preliminary results,'' Ear Hear. **15**, 210–223.

Smurzynski, J., Jung, M. D., Lafreniere, D., Kim, D. O., Kamath, V., Rowe, J., Holman, M. C., and Leonard, G. (**1993**). ''Distortion-product and click-evoked otoacoustic emissions of preterm and full-term infants,'' Ear Hear. **14**, 258–274.

Spektor, A., Leonard, G., Kim, D. O., Jung, M. D., and Smurzynski, J. (**1991**). ''Otoacoustic emissions in normal and hearing-impaired children and normal adults,'' Laryngoscope **101**, 965–976.

Stover, L., and Norton, S. (**1993**). ''The effects of aging on otoacoustic emissions,'' J. Acoust. Soc. Am. **94**, 2670–2681.

Wier, C. C., Pasanen, E. G., and McFadden, D. (**1988**). ''Partial dissociation of spontaneous otoacoustic emissions and distortion products during aspirin use in humans,'' J. Acoust. Soc. Am. **84**, 230–237.

Zizz, C. A., and Glattke, T. J. (**1988**). ''Reliability of spontaneous otoacoustic emission tuning curve measures,'' J. Speech Hear. Res. **31**, 616–619.

Zwicker, E. (**1983**). ''Delayed evoked oto-acoustic emissions and their suppression by Gaussian-shaped pressure impulses,'' Hearing Res. **11**, 359–371.

Zwicker, E., and Schloth, E. (**1984**). ''Interrelation of different oto-acoustic emissions,'' J. Acoust. Soc. Am. **75**, 1148–1154.

Measurement of distortion product phase in the ear canal of the cat

P. F. Fahey
Department of Physics, University of Scranton, Scranton, Pennsylvania 18510

Jont B. Allen
AT&T Labs, Florham Park, New Jersey 07932

(Received 11 October 1996; revised 25 June 1997; accepted 26 June 1997)

Amplitudes of odd order distortion products (DPs) that are detected in animal ear canals have been used to probe cochlear health, to search for cochlear amplification, and to measure aspects of cochlear mechanical frequency response. Like the DP amplitude, DP phase is also an important measure of the cochlear mechanical response. Reported here are measurements of DP phase in the ear canal of the cat. The phase data show frequency-dependent time delays. One of these delays is a function of f_2, the frequency of the higher-frequency primary. Hence the DP phase ϕ_d is of the form $\phi_d=\phi_0+\omega_d\tau$, where ω_d is the DP angular frequency and τ is a fixed time delay. Our results show that ϕ_d is independent of input level a_2 as long as the ratio $a_2/a_1\leqslant 2$, where a_2 and a_1 are the amplitudes of the input tones. As a_2/a_1 becomes greater than two, the fixed time delays increase for DPs whose frequencies are less than the frequencies of the input tones. When both levels are varied together the delay increases as the levels decrease. There can be phase changes as large as π associated with deep nulls in the DP magnitude for the two lower-frequency DPs. Features of the nulls may be modeled assuming that there is partial reflection of the DP wave from the DP place. The assumption of energy reemitted from the DP place also explains amplitude-ratio-dependent time delays and 2π level-dependent bifurcations in phase. The DP phase shows different dependencies for $f_2<1$ kHz compared to $f_2>2$ kHz. © *1997 Acoustical Society of America.* [S0001-4966(97)02111-5]

PACS numbers: 43.64.Kc, 43.64.Bt, 43.64.Jb [RDF]

INTRODUCTION

Intermodulation distortion products (DPs) that are generated by the nonlinear mechanical response of the inner ear to an input of two primary tones are detectable in the ear canal. Since mammals, birds, and reptiles have nonlinear inner ears and, since DPs generally show good signal to noise ratios, DPs can be a useful noninvasive probe of the mechanical response of the cochlea. Indeed, attempts have been made to use DPs as an objective measure of hearing integrity in humans (measurement issues summarized in Whitehead *et al.*, 1994); as a probe of cochlear amplification (Allen and Fahey, 1992); and as a probe of the cochlear frequency response (Brown and Gaskill, 1990; Brown *et al.*, 1992; Allen and Fahey, 1993). Reported here are measurements in cat of the phase of the odd order DPs at frequency f_d equal to $2f_1-f_2$, $3f_1-2f_2$, $2f_2-f_1$, and $3f_2-2f_1$ (f_1 is the frequency of the lower-frequency primary whose respective amplitude is a_1 and f_2 is the frequency of the higher-frequency primary whose respective amplitude is a_2).

Frequencies of the odd order DPs are given by the general expression $f_d=f_1+n(f_2-f_1)=f_1+n\Delta f$, where n is either a positive or a negative integer. If $n\geqslant 2$, the DPs have frequencies greater than either of the primary frequencies and if $n\leqslant -1$ the DP frequency is less than either of the primary frequencies. (When $n=0$, 1 the equation above evaluates to one of the primary frequencies.) Because, when reading the text, it is easy to confuse $f_d=2f_1-f_2$ and $f_d=2f_2-f_1$, we will use the following notation throughout the text. $f_d(-2)=f_1-2\Delta f=3f_1-2f_2$, $f_d(-1)=f_1-\Delta f=2f_1-f2$, $f_d(+2)=f_1+2\Delta f=2f_2-f_1$, and $f_d(+3)=3f_2-2f_1$. $f_d(-1)$ and $f_d(+2)$ are said to be third order or cubic DPs and $f_d(-2)$ and $f_d(+3)$ are said to be fifth order. In a power series expansion, third-order DPs are generated by any odd power in the expansion that is cubic or greater. Fifth-order DPs are generated by any odd power in the expansion that is fifth order or greater. The notation that we suggest identifies the DPs as though they are components in a spectrum centered about the primaries. A negative number in the parentheses means the DP is lower frequency than the primaries and a positive number in the parentheses means that the DP has a higher frequency than either primary. The number in the parentheses is the same as the coefficient of f_2 in the DP.

I. DISTORTION PRODUCT MEASUREMENT

A. Animal preparation and sound system

Both the sound system and the animal preparations in this study were essentially the same as in Allen (1983), except that there was no surgery to expose the auditory nerve. The animals were anesthetized with sodium pentobarbitol and the bulla and septum were open. Enough of the ear canal was removed to allow positioning of the probe microphone to within a few millimeters of the tympanic membrane. Input tone bursts of 40-ms duration and 50% duty cycle were digitally synthesized. The DP phase is the fast Fourier transform (FFT) phase of the time response waveform built from ten

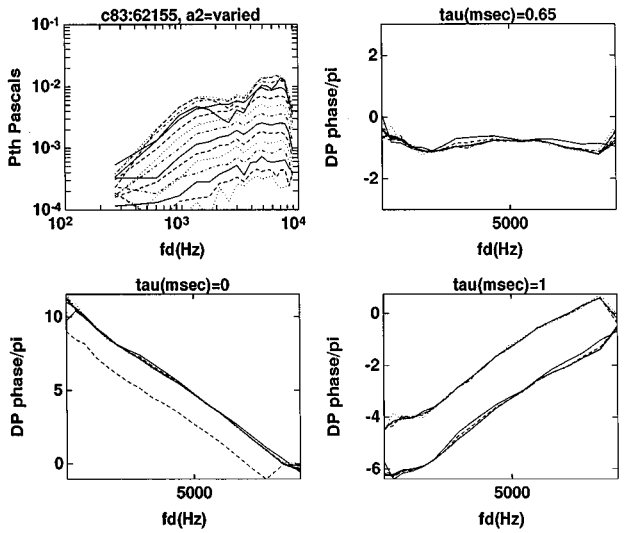


FIG. 1. The top left panel shows the magnitude of the $f_d(-1)=2f_1-f_2$ DP as a function of DP frequency for constant $f_2=9$ kHz. Each magnitude curve was measured at a level wherein a_2 was decreased by 3 dB from the previous curve while a_1 was held constant at 1.0 Pa (1.0 Pa is 94 dB SPL). The DP amplitude was maximum when the value of a_2 was 6 dB less than the value of a_1. The value of the phase was independent of the level of a_2 for $52 \leqslant a_2 \leqslant 85$ dB SPL. The bottom left panel shows the phase of the DP as a function of DP frequency for no subtracted fixed time delay. In the top right panel the phase is plotted after a fixed time delay of 0.65 ms has been subtracted and in the bottom right panel phase is plotted after subtracting a fixed delay of 1.0 ms. It is evident that a fixed delay of 0.65 ms does the best job of making the phase independent of frequency. The frequency axis in the magnitude plot is logarithmic while the frequency axes in the phase plots are linear.

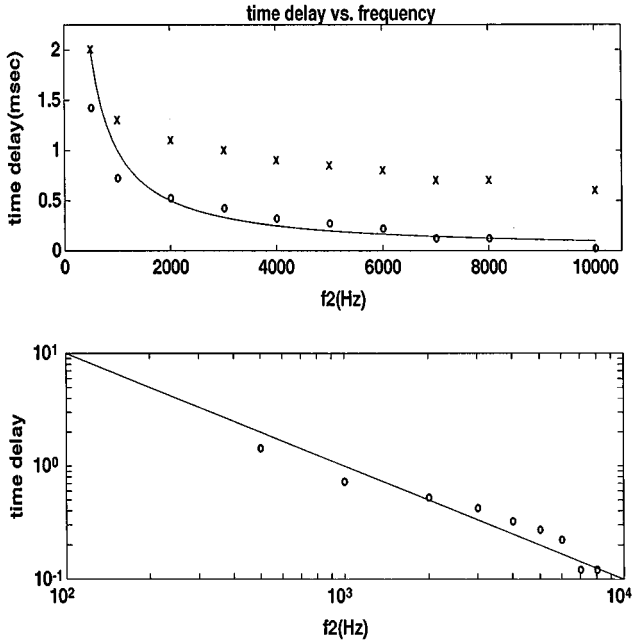


FIG. 2. The top plot shows the values of the fixed time delays τ plotted versus f_2, symbolized by $\times$'s; $\tau-0.6$ ms$=\Delta\tau$ is represented by the $\bigcirc$'s; and the curve is a plot of $1/f_2$ vs f_2. A reasonable first approximation to $\Delta\tau$ is $1/f_2$ as shown by the solid curve. The bottom plot shows the logarithm of $\Delta\tau$ versus the logarithm of frequency. The $\bigcirc$'s are the data and the solid line is $1/f_2$.

averages of the response. For the distortion products reported here the response was not sensitive to the tone burst windowing, whether Kaiser, raised cosine, or none. The data are typical across more than a dozen animals.

B. The measurement

For each animal studied, both the sound delivery system impedance Z_0 and the input impedance of the animal preparation Z_{in} were measured after Allen (1986) and Voss and Allen (1994). The DP pressure measured in the ear canal P_{ec} was expressed as a Thevenin equivalent source pressure P_{Th} given by $P_{\text{Th}}=P_{\text{ec}}(1+Z_{\text{in}}/Z_0)$ (Fahey and Allen, 1986, 1988). Since DP volume velocity in the ear canal due to the DP pressure source in the cochlea depends upon the value of the transducer impedance, the use of the Thevenin equivalent pressure reduces the experimental variations of the results due to the transducer impedance. For most of the data presented here for $0.8\text{ kHz}<f_2<8\text{ kHz}$, $P_{\text{Th}}\approx 2P_{\text{ec}}$ since $Z_0 \approx Z_{\text{in}}$. Therefore while the measurements presented here are of P_{Th} phase, they are similar to raw P_{ec} phase. The magnitude and phase of the DPs were measured by taking the FFT of the time waveform of the ear canal pressure. Since the phase has an uncertainty of 2π, when plotting phase versus frequency, the phase data was unwrapped so that change in the phase from one point to the next was minimized. By first subtracting that part of the phase that is due to a fixed time delay we eliminate much of the potential for uncertainty in the phase unwrapping. In the data that follows it will occasionally be useful to keep this uncertainty in mind when viewing some of the features in the plotted phases. Stover *et al.* (1996) used DP magnitude and phase to define a ''filter function'' and its inverse Fourier transform that was called the ''IFFT waveform.'' We will refer to these constructions as the FILTF and WAVEF. We constructed FILTF's by inserting a 128 point spline fit of the data into a spectrum that was padded with zeroes both at frequencies below the lowest value of f_d and at frequencies above the highest value of f_d. A 1024 point WAVEF was generated from the FILTF inverse Fourier transform. The functional shape of the WAVEF resembles a system impulse response, but the WAVEF is not a system impulse response because it is nonlinear in origin. The WAVEF is a different way of presenting the same data as the DP magnitude and phase data. The advantage of the WAVEF presentation of the data is that delays that are implicit in the raw phase data are presented weighted by the energy at the various delays. This technique

TABLE I. Fixed time delay as a function of f_2.

f_2 (kHz)	0.5	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	10.0	12.0	17.0
τ (ms)	2.0	1.3	1.1	1.0	0.9	0.85	0.8	0.7	0.7	0.60	0.60	0.60
$\Delta\tau=\tau-0.6$	1.4	0.7	0.5	0.4	0.3	0.25	0.2	0.1	0.1	⋯	⋯	⋯
$f_2\Delta\tau$	0.7	0.7	1.0	1.2	1.2	1.25	1.2	0.7	0.8	⋯	⋯	⋯

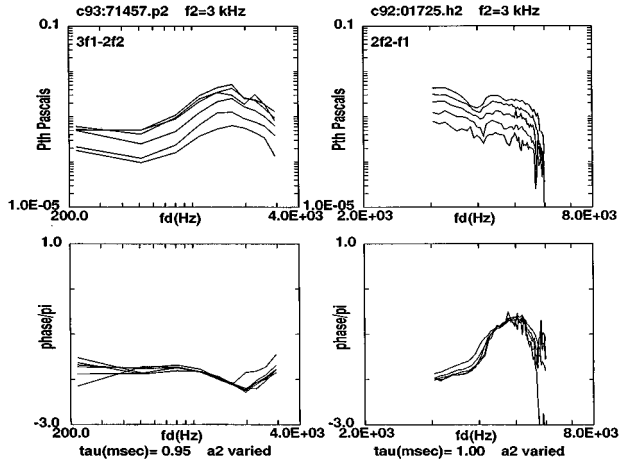


FIG. 3. In the left panels magnitude and phase are plotted as a function of f_d for the lower frequency fifth-order DP, $f_d(-2)=3f_1-2f_2$, as a_2 is decremented from $a_2=0.78$ Pa in 3-dB steps at a constant $a_1=0.77$ Pa. The lowest value of a_2 is 0.14 Pa. The phase changes little with a_2 if $a_2<0.5a_1$. In the right panels the higher-frequency third-order DP, $f_d(+2)=2f_2-f_1$, magnitude, and phase are plotted as functions of DP frequency. The value of $a_1=0.65$ Pa and $0.16\leqslant a_2\leqslant 0.65$ Pa.

is conceptually more mathematically simple than the method of Brown *et al.* (1996) and it has the added potential of giving more information about signal delays.

C. Data presentation

Figure 1 displays the effects of subtracting a fixed time delay from the phase data and Fig. 2 is a plot of Table I. Then Figs. 3–8 have the same presentation format. Each is a plot of the ear canal Thevenin equivalent DP pressure plotted against the DP frequency f_d. Each curve is measured at a different level of the input. At the top of the figure are listed the animal and file and the constant value of f_2. At the bottom of the figure are the value of the fixed time delay τ in ms. That part of the phase that is due to the fixed time delay

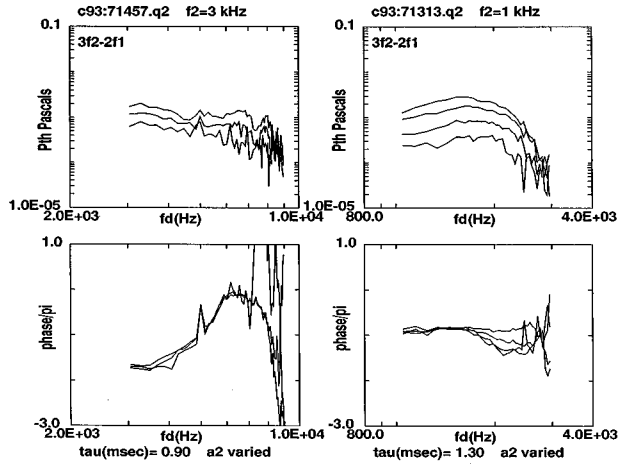


FIG. 4. The higher-frequency fifth-order DP, $f_d(+3)$, magnitude, and phase are plotted as functions of DP frequency. a_2 is decremented in 3-dB steps. In the left panels the phase is independent of a_2 for $0.39\leqslant a_2\leqslant 0.78$ Pa while $a_1=0.77$ Pa. In the last curve the data above $f_d=7$ kHz is in the noise. In the right panel, starting at $a_2=0.38$ Pa, a_2 is decremented in 3-dB steps to a final value of $a_2=0.13$ Pa while a_1 is constant at 0.37 Pa. The phase changes little with a_2 when $1<f_d<1.7$ kHz.

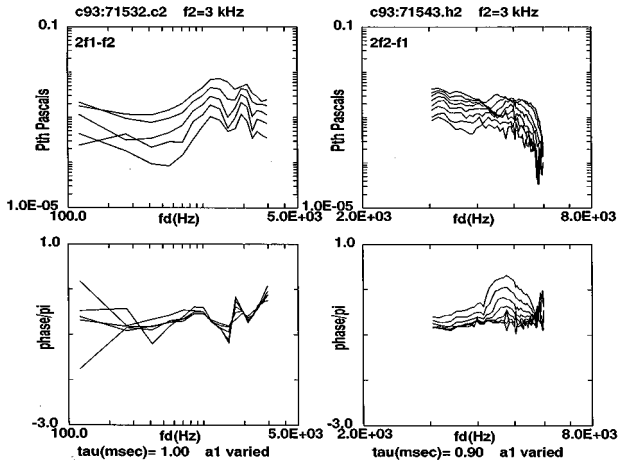


FIG. 5. In the left panels magnitude and phase are plotted as functions of f_d for the lower-frequency third-order DP, $f_d(-1)$, as a_1 is decremented in 3-dB steps from $a_1=0.63$ Pa to $a_1=0.15$ Pa at a constant $a_2=0.63$ Pa. The phase changes little with a_1 except in the vicinity of the nulls in the magnitude between 1.5 and 2.5 kHz. In the right panels magnitude and phase are plotted as a function of f_d for the higher-frequency third-order DP, $f_d(+2)$, as a_1 is decremented in 3-dB steps from $a_1=0.75$ Pa to $a_1=0.067$ Pa at a constant $a_2=0.78$ Pa. The lowest four phase curves are independent of a_1 level.

$\omega_d\tau$ has been subtracted from the phase that is plotted. Therefore given that $\phi_d=\phi_0+\omega_d\tau$, the phase that is plotted is ϕ_0. Also identified at the bottom of the plot is the level that was varied. The data in this study are at fixed values of f_2. Here, f_d is varied by changing the value of f_1.

The input levels used in this study were as high as 1.0 Pa (or 94 dB SPL) and were higher than levels generally used in human studies. Correspondingly, the DP levels were as high as 0.02 Pa (or 60 dB SPL).

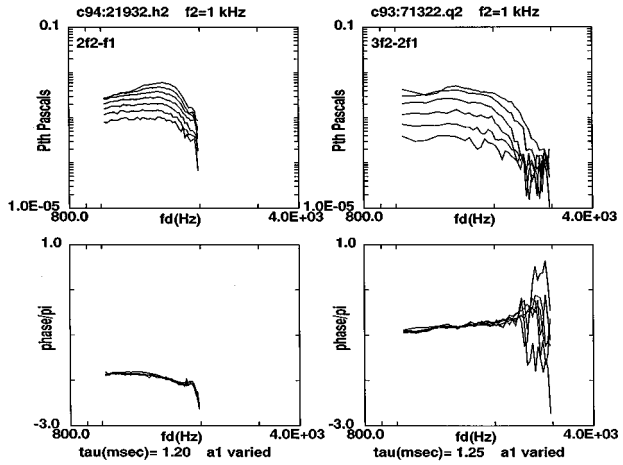


FIG. 6. In the left panels magnitude and phase are plotted as functions of f_d for the higher-frequency third-order DP, $f_d(+2)$, as a_1 is decremented in 3-dB steps starting at $a_1=0.73$ Pa and ending at $a_1=0.09$ Pa at a constant $a_2=0.74$ Pa. All seven phase curves superimpose. In the right panels magnitude and phase as functions of f_d are shown for the higher-frequency fifth-order DP, $f_d(+3)$, as a_1 is varied from $a_1=0.62$ Pa to $a_1=0.11$ Pa and a_2 is constant at 0.64 Pa. Notice that the phase does not depend on a_1 at this frequency of f_2. For $f_d>2.3$ kHz the magnitude is in the noise at the lower values of a_1; hence, the phase is noisy also. When $f_2\leqslant 1.0$ kHz the phases of the higher-frequency DPs seem to show less dependence on a_1 than at the higher f_2.

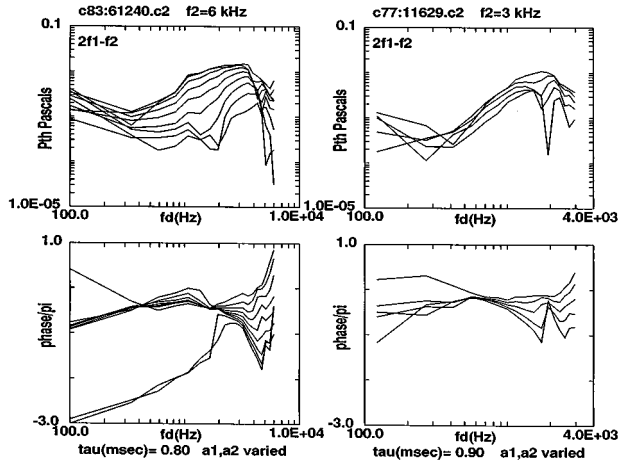


FIG. 7. Magnitude and phase are plotted as a function of f_d for the lower-frequency third-order DP, $f_d(-1)$, as a_1 and a_2 are decremented in 3-dB steps beginning at $a_1=a_2=0.95$ Pa and ending at $a_1=a_2=0.06$ Pa in the left panel and from 1.0 to 0.26 Pa in the right panel. In the left panels at a value of f_d just less than 2 kHz the phase seems to be independent of level (except for the phase curves measured at the two lowest levels). For higher frequencies than this fixed phase point the phase appears as though it rotates about this point in a clockwise direction, i.e., the slope of the phase becomes more negative as the level decreases. At frequencies less than the fixed point the phase varies slightly with decreasing level. For the two phase curves at the two lowest input levels a π phase shift occurs at about this frequency. Comparing the two lowest input level phase curves with the higher level curves shows that there is a level-dependent phase bifurcation of 2π for $f_d<2$ kHz. In the right panels, while there is not a fixed phase point as in the previous figure, the phase varies little around $f_d=0.7$ kHz. In this figure, at the lowest level of the primaries a sharp null is evident in the magnitude at $f_d=2$ kHz. There is also a π phase shift at this frequency.

II. RESULTS

A. Time delays

In Fig. 1 the ear canal Thevenin equivalent pressure magnitude of a $f_d(-1)=2f_1-f_2$ DP is plotted versus its own frequency in the top left panel; the phase versus frequency with no fixed delay subtracted is plotted in the bottom left panel; in the top right panel a fixed time delay of 0.65 ms has been subtracted from the phase; and, in the bottom right panel a fixed delay of 1.0 ms has been subtracted. Here $f_2=9$ kHz and the subtracted time delay of 0.65 ms removes most of the phase variation at low frequencies (and, in this file, also at high frequencies).

The modeling results of Matthews and Molnar (1986) show that most of the DP energy is generated in the vicinity of the f_2 place. There are also good general theoretical reasons that DPs generated by a saturating nonlinearity would be generated near the f_2 places. The maximum DP is generated when the levels of the two inputs to the nonlinearity are approximately equal. For cochlear excitations that region is near the f_2 place. The data presented in this study are at constant f_2.

In Table I we tabulate the values of the fixed time delay as a function of f_2 under the condition that one of the primary levels was constant and the other primary level was varied. For the $f_d(-1)$ and $f_d(-2)$ DPs a further condition is that $a_2/a_1\leqslant 2$. Under these conditions the phase is independent of the level of the varied primary and the low-frequency slope of the phase as a function of $\omega_d(=2\pi f_d)$ is constant; therefore, τ is constant. As f_2 increases the time delay decreases. These values are a good representation of the fixed time delay across the dozen animals used in this study and across the four DPs. For higher values of f_2 the time delays come primarily from $f_d(-1)$ and $f_d(-2)$ data records. At values of f_2 on the order of 10 kHz or greater, the fixed time delay reaches an asymptote of about 0.6 ms. The third line of Table I shows the values of the time delays with the value of the asymptotic delay subtracted and the fourth line shows the values in the third line multiplied by their respective f_2. This product is of order one, suggesting that the time delays after the asymptotic delay has been subtracted is, to a first approximation, equal to $1/f_2$. A graph of the results in the table is Fig. 2.

When a_2/a_1 is greater than about two or three, the measured time delays for the $f_d(-1)$ and the $f_d(-2)$ DPs can be greater than the values in Table I. The increase ranged from 1.0 ms at lower values of f_2 to 0.3 ms at the highest values. The effect of the ratio a_2/a_1 is clearly seen in Fig. 9 and will be discussed in a following section. The longer delays could well be due to an increase in energy that arrives later because it has been reflected from the DP place.

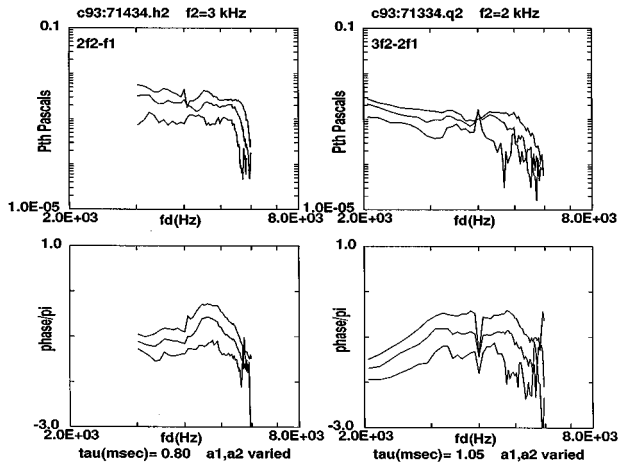


FIG. 8. a_1 and a_2 are decremented together by 3 dB each for the $f_d(+2)$ DP in the left panels and for the $f_d(+3)$ DP in the right panels. For both of these DPs the ratio $a_2/a_1=1.0$ and the initial values are $a_1=a_2=0.7$ Pa and the final values are 0.35 Pa. The sharp null in the phase at $f_d=4$ kHz in the right panels is probably due to mixing with the $2f_1+f_2$ DP which moves down in frequency as $f_d(+3)$ moves up (with decreasing f_1).

B. Phase dependence on primary level(s)

1. a_1 constant, a_2 varied

Figures 1, 3, and 4 show both DP magnitude and phase as a_1 is held constant and a_2 is decremented in 3-dB steps. In Fig. 1 the $f_d(-1)$ DP phase data shows that, except for the data acquired at the four highest levels of a_2, the phase curves are independent of level. In Fig. 3 (left) the same general independence of phase with level is shown for the $f_d(-2)$ DP. Phase as a function of a_2 and f_d is is shown in Fig. 3 (right) for the $f_d(+2)$ DP and again there is independence of a_2. Figure 4 (left) illustrates the independence of phase with a_2 level for the $f_d(+3)$ DP. This behavior is typical for $f_2\geqslant 2.0$ kHz. However, Fig. 4 (right) shows that

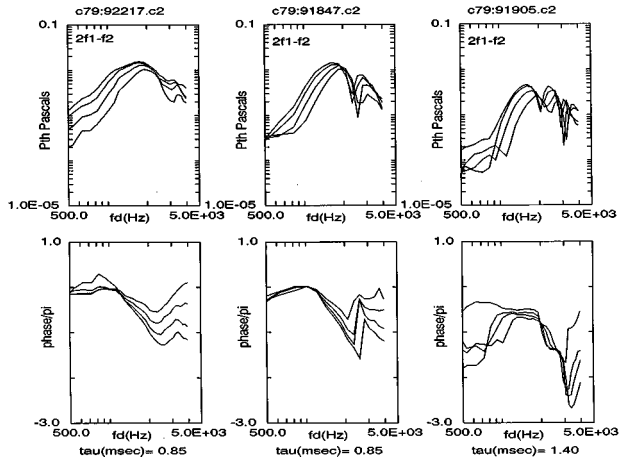


FIG. 9. In this figure a_1 and a_2 are decremented together in 3-dB steps while the ratio a_2/a_1 changes from 0.3 in the left panels to 1.0 in the middle panels to 3.0 in the right panels. The density of nulls increases as the ratio a_2/a_1 increases. In the left panel $0.31 \leqslant a_1 \leqslant 0.89$ Pa and $0.09 \leqslant a_2 \leqslant 0.34$ Pa. In the middle panel $0.22 \leqslant a_1 = a_2 \leqslant 0.61$ Pa. In the right panel $0.11 \leqslant a_1 \leqslant 0.31$ Pa and $0.33 \leqslant a_2 \leqslant 0.92$ Pa.

for the $f_d(+3)$ DP when $f_2 \leqslant 1.0$ kHz there is some independence of phase with changing a_2, but only for f_d within an octave or so of f_2.

2. a_1 varied, a_2 constant

The marked independence of phase across the DP frequency range with changing a_2 is not generally as pronounced with changing a_1. Figure 5 (left) shows that, as a_1 is decreased by 3 dB, the phase changes in the vicinity of nulls in the $f_d(-1)$ DP magnitude versus frequency curve. Indeed, the amplitude peak to valley ratio of the nulls seems to increase as a_1 is decremented. We will show in a later figure that the presence of nulls increases as the a_2/a_1 ratio increases.

For the $f_d(+2)$ DP when $f_2 \geqslant 2.0$ kHz we observe that the phase is independent of a_1 only at the lowest levels [as seen in Fig. 5 (right)]. This is also seen in the $f_d(+3)$ data. When $f_2 \leqslant 1.0$ kHz, the phase of the $f_d(+2)$ DP is generally independent of a_1 as is shown in Fig. 6 (left). Likewise, when the DP is $f_d(+3)$ and $f_2 \leqslant 1.0$ kHz, we also observe independence of phase with a_1 as shown in Fig. 6 (right).

Comparing the data of this and the previous section with the data of the next section, it is evident that phase changes much less when only one primary level is varied than when both primary levels are varied together.

C. Both a_1 and a_2 decreased

Figure 7 shows $f_d(-1)$ DP phases and magnitudes as both a_1 and a_2 are each decremented by 3 dB. Qualitatively, the same phase dependence on level is observed in the $f_d(-2)$ phase (e.g., see phase in Fig. 12). In Fig. 7, at $f_d \approx 2$ kHz (except for the two phase curves measured at the two lowest levels) the phase seems to be almost independent of input level and, as the levels decrease, the phase above $f_d = 2$ kHz decreases in a regular way. This pattern is commonly observed when changing a_1 and a_2 together. When f_d is greater than the value of f_d where the magnitude is maximum, $f_d^{\max}$, the phase decreases when both levels are decremented together. It is also observed that, as a_1 and a_2 are decremented together, the frequencies at which the sharp nulls occur in the DP magnitude increase. For the two higher-frequency DPs the change in phase with the decrease of both levels seems to be a simple decrease in phase as seen in the panels of Fig. 8.

D. Sharp nulls

In Fig. 7 (right panel) there is an example of a sharp null in the DP magnitude at $f_d = 2$ kHz. In the phase a shift of π radians is observed at the null. When nulls begin to form as the level is decreased, the phase always shows a discontinuity at the null frequency. For the deepest nulls the change in phase is π radians. The sharpness of the nulls and the sharpness of the π phase shift are indicative of wave interference.

The appearance of nulls in the data is a function of the ratio a_2/a_1. When $a_2/a_1 > 2$ the density of nulls is greater. Figure 9 illustrates this effect. In each set of panels both primaries are decremented together, while the ratio a_2/a_1 changes from 0.3 in the left panels to 1.0 in the middle panels to 3.0 in the right panels. The nulling is least when the ratio is least. When the data of Fig. 9 (right) is plotted on a linear frequency scale (not shown) it is evident that the frequency spacing between the nulls is approximately constant (possibly increasing slightly as f_d decreases). When both a_1 and a_2 are varied together the nulls shift toward higher values of f_d as the primary levels decrease.

III. DISCUSSION

A. Time delays

As previously mentioned, the fixed time delays in Table I were chosen to minimize the slope of the phase for $f_d \leqslant f_d^{\max}$. We also measured time delays by taking the negative of the slope of the complete phase versus linear angular frequency curve (Allen, 1983; Kimberley *et al.*, 1993; O Mahoney and Kemp, 1995) and, within the error of both techniques, measured the same values for delay. The correspondence of the two methods is clear from the phase plots of Fig. 1. The negative of the slope of the phase in the lower left-hand panel is equal to 0.65 ms. This is the delay that has been subtracted from the lower left panel phase in the plot of the upper right-hand panel.

Furthermore, following Stover *et al.* (1996) we built FILTF's from our data that we inverse Fourier transformed to get WAVEF's. As observed by them, the maximum of the magnitude of the WAVEF can be used to define a latency. The latencies found using this technique, with $a_2/a_1 < 1$, gave us the same values as shown in Table I, within the error of the techniques. Of course, this is to be expected from the mathematical properties of the inverse Fourier transform if, globally, the phase slope is approximately constant. The top panels of Fig. 10 show the WAVEF magnitudes for the same data as shown in Fig. 1 and in Fig. 5 (left). The peak in the magnitude occurs at times consistent with the values of the delay for $f_2 = 3$ kHz and $f_2 = 9$ kHz, respectively. The time delays in Table I characterize all of the animals used in this

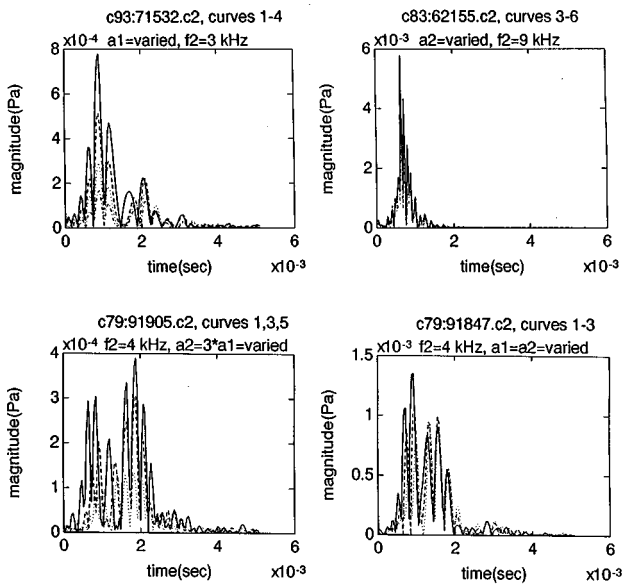


FIG. 10. WAVEF's constructed after Stover *et al.* (1996) have been constructed for some of the data in Fig. 5 (top left panel) and in Fig. 1 (top right panel). The magnitude of the responses is maximum at times that coincide with the delays determined by subtracting a fixed delay. The amplitudes of the impulse responses scale as expected from the curves in Fig. 5 (left) and Fig. 1. Notice that in the top left panel a second wave at a later time ($\approx$2 ms) grows relative to the first wave as a_1 decreases. The bottom panels of this figure show WAVEF magnitudes constructed from the data of the right and middle panels of Fig. 9. In the bottom left panel of this figure are the data taken when $a2/a1=3.0$ and in the bottom right panel when $a2/a1=1.0$. In the bottom left panel the amplitude of the later peak in the magnitude is larger than at the earlier peak. This is why the fixed delay as seen in Fig. 9 (right) is greater than in the Fig. 9 left or middle phase panels.

study. The values in Table I apply to every animal that we studied to within ± 0.3 ms, except at the lowest value of f_2, where the range was ± 0.5 ms. Kimberley *et al.* (1993) have shown that the time delay (also called latency) in humans decreases as the level of both primaries are increased together. This is also apparent in the cat data of Figs. 7 and 9. For example, calculating the delay from the complete phase versus linear frequency curve for Fig. 7 (left), the slope increases by a factor of about 1.5 from the highest levels to lowest.

The delays published here for a cat are much smaller than the delays in a human in the same frequency range. In a human, when $f_2 \approx 4$ kHz, the latency is on the order of 3 to 4 ms (Brown *et al.*, 1992; Kimberley *et al.*, 1993; Whitehead *et al.*, 1996; Stover *et al.*, 1996). Stover *et al.* have observed that the latency for some other odd order DPs is the same as for the $f_d(-1)$ DP. The cat data in this study also suggests that the fixed time delay is independent of DP order. This is certainly not surprising if it is assumed that the various odd order DPs are generated by the same mechanism at the same place. Whitehead *et al.* (1996) directly measured the time to onset of the DP signal in the ear canal and found times consistent with the measurement of latency from the slope of the phase. Stover *et al.* (1996) have shown that the delays derived from the WAVEF are commensurate with latency determined by the slope of the phase. Hence the four methods used to measure delay all yield consistent results. The phase data that we show in this study has the fixed delay subtracted from the total phase so that we emphasize and magnify the other features in the phase data (i.e., the nulls and what is happening to the phase above $f_d = f_d^{\max}$).

When the DP delays are compared with neural delays (of inputs at $f=f_d$ at the f_2 place) in the same kind of animal preparation (Allen, 1983), the DP delays match up well with minimum values for neural delays (after subtracting out approximately 0.9 ms of synaptic delay. These delays also generally correspond to the latencies of neural onset to condensation clicks in a cat of Kim and Molnar (1979, Fig. 8), although for characteristic frequencies greater than 2.0 kHz their latencies are systematically smaller than our fixed time delays. Part of neural delay is due to one-way travel time. Initially, one might think that some of the DP delay as due to a two-way travel time (for fixed f_2), i.e., the travel time of f_1 to the f_2 place and the travel time out of the cochlea of the DP. However, the travel time of f_1 to the f_2 place changes little as f_1 decreases; hence, the DP phase delay is primarily due to travel time of the DP excitation to the ear canal. That the minimum neural delay and the DP delay are approximately the same as a function of frequency again suggests that the DP is generated primarily at the f_2 place.

In a human, there seems to be some question on whether the DP delay is a measure of two-way travel time or one-way travel time (Kimberley *et al.*, 1993; O Mahoney and Kemp, 1995; Bowman *et al.*, 1997). Assuming that human DP biophysics is the same as a cat's, and that the cat data show that neural travel time delay and DP delay are the same, it is clear that, for at least one DP component, the travel delay is due to one way travel time.

The relationship between fixed delay and f_2 is both what one would expect from models of cochlear macromechanics and from the theory of causal transfer functions. From Table I it is evident that there seems to be a minimum delay that is frequency independent and would be due to nonspecific delays such as measurement system delay (determined from the impulses response as derived from the wideband frequency response of the animal preparation) of between 0.2 and 0.3 ms and delay in communicating a signal from the base of the cochlea to the ear canal. This minimum delay is approximately 0.6 ms. If this value is subtracted from all of the other delay values and then the answer is multiplied by the respective f_2 for $1.0 \leqslant f_2 \leqslant 8.0$ kHz, this time-frequency product is a constant N of order 1 (specifically, $0.7 \leqslant N \leqslant 1.2$), showing that the low-frequency limit is basically proportional to $1/f_2$ (Fig. 2). The correlation of delay with f_2 can be expected if it is assumed that DP generation is at the f_2 place. For direct excitation of the cochlea, the proportionality of delay to $1/f_{\mathrm{CF}}$ (which is the same as f_2 in this context) for low-frequency excitation has been derived by Allen (1983) using the WKB-approximation cochlear model. Furthermore, using one of the relations between the real part and the imaginary part of a general transfer function (or a general immittance) derived by Bode (Bode, 1945, Chap. 13) it can be shown that the time delay in the low-frequency limit is inversely proportional to the resonant frequency (or the band-edge frequency for a low-pass system). The constant of proportionality depends only on the shape of the transfer function. (This relationship is easily seen in the expression for phase of a

damped simple harmonic oscillator where the constant of proportionality is the inverse of $2\pi Q$, where Q is the quality factor.) Assuming that the shape of the transfer function varies little with f_2; then, the proportionality of τ to $1/f_2$ would be expected.

B. The level dependence

Both the magnitude and phase curves show that there is much more regularity to the DP generation when only one input level is varied than when both levels are varied (e.g., Zwicker, 1981; Zwicker and Harris, 1990; Fahey and Allen, 1986, 1988; Whitehead *et al.*, 1995).

For the $f_d(-1)$ DP and the $f_d(-2)$ DP, when $a_2 < (a_1 = \text{const})$, the phase and the shape of the magnitude curve are invariant with respect to the value of a_2, as seen in the example of Fig. 1 and of Fig. 3 (left). Given that basilar membrane (BM) response changes shape (Rhode, 1980; Ruggero and Rich, 1991) with level, it could be expected that the DP generation could not be invariant with level. However, the conditions of the DP generation reported here, when thought of in terms of a two-tone suppression experiment where f_1 is a low-frequency suppressor at the f_2 place, could decrease the sensitivity of the most sensitive response region (or ''tip'') of the BM and/or the neural frequency tuning response at the f_2 place. The levels used were generally high, as the varied primary usually ranged between roughly 94 and 54 dB SPL. The primary that was not varied was usually fixed somewhere between 94 and 84 dB SPL. For the $f_d(-1)$ DP with a_2 varied the value of the fixed a_1 was in the range where it would suppress the a_2 response at its characteristic place. Basically, the most input sensitive region of the response would have been removed (Fahey and Allen, 1985; Ruggero *et al.*, 1992).

Given the measurements of magnitude and phase for the a_1 constant with $a_2/a_1 < 2$ for the two lower-frequency DPs, the data show that the DP generation process simply scales the DP magnitude with a_2 and that the frequency features of the DP are invariant under a_2 scaling.

When a_1 is varied with a_2 fixed, the phase is more dependent upon input level for the two lower-frequency DPs. In Fig. 5 (left) the phase varies around the frequencies where there are nulls in the the magnitude response curve. The phase change upon stepping through the null frequency gets greater as a_1 decreases. The WAVEF of the data of Fig. 5 (left panel) is shown in the top left panel of Fig. 10. The energy seems greatest at a time delay of ≈ 1 ms: however, there is also a second peak in the energy at ≈ 2 ms. As a_1 decreases (and the depth of the nulls increases) the peak in the energy at ≈ 2 ms increases relative to the peak at ≈ 1 ms. This same phase behavior is seen in the $f_d(-2)$ phase data (not shown). When a_1 only is varied, the nulls in the magnitude response do not change frequency.

For both the two higher-frequency DPs and for the lower-frequency DPs as long as $a_2/a_1 < 2$ phase is generally independent of the changing of one input level while the other input level is held constant.

C. Effect of the a_2/a_1 ratio as both levels are varied

In Fig. 9 we show that the presence of nulls depends upon the ratio a_2/a_1. When this ratio is large there are more nulls and/or the nulls are more evident. Also, from the values of the subtracted time delays given at the bottom of each phase panel, the time delay is largest at the largest value of the ratio. When both of the primary levels are varied together, the nulls in the magnitude of the $f_d(-1)$ and $f_d(-2)$ DPs shift to higher values of f_d as the levels decrease. The shifting of human DP nulls with level has been reported by He and Schmiedt (1993). They also show that the frequency spacing of nulls in a human is approximately 1/10 of an octave. In Fig. 9 (right), using both the magnitude and phase information it appears that there are nulls at $f_d = 1$, 2, and 3 kHz for an $f_2 = 4$ kHz. If the magnitude data in Fig. 9 (right) are replotted on a linear frequency scale (not shown), it is evident that the frequency spacing between the adjacent nulls is, roughly, constant.

WAVEF's of the data at $a_2/a_1 = 3$ and $a_2/a_1 = 1$ are shown in the bottom panels of Fig. 10. We find that the energy is more spread over time and peaks at later times as the a_2/a_1 ratio increases. Under some conditions the energy that arrives later can be greater than the energy that arrives earlier. The bottom panels of Fig. 10 also show that the time of the first peak increases as both of the primary levels decrease. This is consistent with the nulls moving to higher frequency with decreasing level and with the latency increasing as primary levels decrease (Kimberley *et al.*, 1993; Stover *et al.*, 1996; Whitehead *et al.*, 1996).

The relative growth of the later peaks to the early peak with decreasing both a_1 and a_2 that is evident in the WAVEF presented here for a cat was observed by Stover *et al.* (1996) in a human.

The number of nulls per Hz (the null density) in a human (He and Schmiedt, 1993) is much greater than the null density in a cat (this study). Another difference between the cat data and the human data is the level of the DPs. In He and Schmiedt the DP level is generally between -15 and 0 dB SPL, while the data of our Fig. 9 is around 30–40 dB SPL. The input levels in the human are also corresponding smaller. The noise floor in our measurements was dependent on frequency, but for $1\text{ kHz} < f_d < 8\text{ kHz}$ it was roughly a constant 10 dB SPL and we did measure DPs down to this level. At these lower levels (data not shown) we found no evidence that the null density would substantially increase as levels decrease. Even at the lowest levels, the null density was not much greater than 2 nulls per octave in a cat. Hence the null density difference between a human and a cat is not due to level differences. The ratio of the null density in the two species is roughly equal to the inverse of the ratio of the latencies. That the null spacing would be inversely proportional to the latency would suggest that propagation delays are implicated in the wave interference producing the nulls.

For a given pair of input levels, the nulls in human $f_d(-1)$ and $f_d(-2)$ data do not occur at the same f_d frequencies; however, the spacing between nulls (or peaks) is similar (Brown *et al.*, 1993b, Fig. 2). In a cat, it is also generally observed that the nulls of the two DPs do not occur at

the same value of f_d, nor do they occur at the same value of f_1.

The higher-frequency DPs, $f_d(+2)$ and $f_d(+3)$, do not show nulls like the two lower frequency DPs. At low levels of the DPs (20 dB SPL or less) we do occasionally observe nulls. However, these nulls are generally neither stable nor regular in frequency as input levels are decreased. The higher-frequency DPs do show predictable frequency-dependent features that are due to mixing with other components, such as the third harmonic of f_1 or the DPs at $2f_1+f_2$ or even $5f_1-2f_2$.

D. Source of the nulls

If it is assumed that the nulls are due to wave interference, then there are two or more waves mixing at the site of the microphone in the ear canal. If there are only two dominant waves, then, for a deep null the phase difference between the two waves of comparable amplitude must be $\approx\pi$. If there are several waves, then the phasor diagram must (almost) close. Waves of differing phases could be due to: (a) a path/time differences, such as path differences due to reflections; (b) a ''birefringence,'' i.e., two (or more) different values for wave velocity (e.g., Hubbard, 1993); or (c) two (or more) different nonlinear mechanisms with different phase at output.

The nulls observed psychophysically (i.e., within the cochlea) in the $f_d(-1)$ DP have been attributed by Zwicker (1981) to the interference of wavelets (i.e., waves from multiple sources).

Parallel approaches to modeling psychophysical DP nulls have been proposed to modeling ear canal DP nulls. Substantial discussion has been devoted in the literature to two possible mechanisms responsible for the microstructure observed in human DP emissions. The one proposed mechanism is the mixing of wavelets (Sun *et al.*, 1994; He and Schmiedt, 1996) and the other mixing due to energy reflected from the DP place (Talmadge *et al.*, 1996; Talmadge *et al.*, 1995; Brown *et al.*, 1996). If we assume that the nulls measured in a cat are due to the same physics as the microstructure measured in a human, then our cat data strongly support reflection from the DP place over distributed source generation of wavelets.

First, if the mixing of wavelets due to a distributed source of DP waves were the source of the nulls in the $f_d(-1)$ and the $f_d(-2)$ DPs, then one would also see nulls in the $f_d(+2)$ and $f_d(+3)$ DPs.

Second, as described below, using a simple semiempirical model, the assumption of reflection of energy from the DP place explains (a) several features of the nulls; (b) the commonly observed 2π phase splitting seen in DP data; and (c) the long delays in the $f_d(-1)$ and $f_d(-2)$ DP when a_2/a_1 is large.

There are several features of the nulls in our data that can be replicated with a simple physical picture. This picture assumes that the distortion product is generated at the f_2 place and that some of the DP wave is propagated basally (toward the stapes) and the remaining energy is propagated apically (toward the helicotrema). The time to propagate from the f_2 place to the stapes is assumed to be $1/f_2$ as the data in Table I would suggest. Waves that reach the stapes may be partially reflected and partially transmitted. The reflection coefficient at the stapes is r_{st}. Waves that propagate apically may be partially reflected at the DP place (Allen *et al.*, 1995). [This is essentially the same assumption that has been used by Zweig and Shera (1995) to explain the periodicity of evoked stimulus frequency emissions that are observed in the human ear canal.] The reflection coefficient at the DP place is r_{DP}. Two ways to model the phase change of the DP wave between the f_2 place and the DP place are to assume either a scaling symmetric cochlea, e.g., after Zweig and Shera (1995) or a nearly scaling symmetric cochlea, e.g., after Allen (1980). The scaling symmetric cochlea would have a phase change from the stapes to the characteristic place that scales as $\hat{\phi}\ \log(f_d)$. The Allen model can be shown to have a basilar membrane delay between the stapes and the characteristic place f_d that can be approximated by a constant divided by $(f_d/f_{\mathrm{ref}})^{\beta}$ or α/f_d^{β}, where α has units of time and f_d is now a relative frequency that is dimensionless. The model basilar membrane response data of Allen yield $\alpha\approx 0.3$ and $\beta\approx 0.8$. ($\beta=1$ would be a scaling symmetric cochlea.) The time for the DP wave to travel from the f_2 place to the DP place is assumed, to a first approximation, to be the time to travel from the stapes to the DP place, minus the time to travel from the f_2 place to the stapes, $1/f_2$. An interferometer with either a delay after the Allen model or a delay after the Zweig and Shera model produces interference nulls that have a frequency spacing that is similar to that of the data. For the scaling symmetric cochlea assumption, a value of $\hat{\phi}\approx 0.31$ produces interference patterns that look like the cat data. A picture of the interference pattern produced by the scaling symmetric cochlea assumption is shown in the top panel of Fig. 11 and the pattern produced by the Allen basilar membrane model is shown in the bottom panel of Fig. 11. To a first approximation the reflection coefficients are assumed to be independent of amplitude. Details of this physical picture will be published separately. With reflections at the stapes and at the DP place there will be an infinite sum of waves that mix at the detector in the ear canal. The phase differences of the various waves that mix at the detector will depend upon the different travel times and upon any phase changes at reflection. The amplitude differences of the wavelets will depend upon the magnitudes of the reflection coefficients (we will assume that there is no significant attenuation on the basilar membrane). Letting r_{st} become zero, the above picture becomes a simple two-source model. The major consequence of $r_{st}\neq 0$ is to the sharpen the nulls and to broaden the maxima of the interference patters. As r_{st} is allowed to increase, one generates the sharp interfence patterns commonly pictured in optics textbooks (Born and Wolf, 1975, Sec. 7.6).

Our data are consistent with values of $r_{st}\leqslant 0.3$. The null density depends upon travel time assumptions and would be different for a human than for a cat. Indeed, as O Mahoney and Kemp (1995) have pointed out, ''In humans ears tone-burst-evoked emissions commonly have ten or more waves of delay.'' An examination of the data in the paper by Kimberley *et al.* (1993) shows a delay of between six and seven waves in human. Our data in Table I show that cats have one

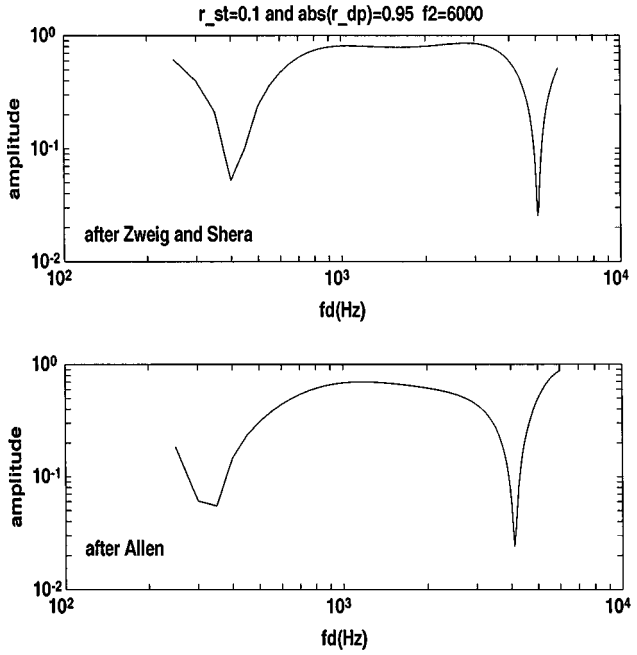


FIG. 11. The top panel shows the pattern of nulls calculated assuming a phase change from the stapes to the characteristic place of a scaling symmetric cochlea after Zweig and Shera (1995). The bottom panel shows the pattern of the nulls assuming the delays from the basilar membrane model of Allen (1980). For each panel, it was assumed that the time for a DP wave to travel from the f_2 place to the stapes was $1/f_2$. The values assumed for r_{st} and r_{DP} are shown at the top of the figure. r_{DP} is complex and its phase was adjusted between the top and the bottom panels in order to position the nulls at convenient values of f_d for the purpose of illustration. In this figure f_2 is assumed to be constant. Similar patterns are produced assuming that f_1 is constant or that f_2/f_1 is constant.

wave of delay. Hence the frequency spacing of such an interference pattern as described above in a human would be somewhere around one-tenth to one-fifth of the cat frequency spacing or from 0.1 to 0.2 kHz. Values of the observed frequency spacing in a human (He and Schmiedt, 1993; Brown *et al.*, 1993a; Piskorski *et al.*, 1995) are in this range.

The null density will depend upon the rate that the phases of the interfering waves changes as the f_d changes. Using the model described above, during a frequency sweep of f_d, we find a maximum of three nulls. This maximum is only slightly dependent on the starting value of f_2. The relative independence on the starting value of f_2 on the maximum number of nulls per f_d sweep is also a feature of the data. The null density in both the model and the data (Piskorski *et al.*, 1995) is the same if f_2 is held constant, if f_1 is held constant, or if the ratio f_1/f_2 is held constant.

How is it possible that more energy arrives later than the energy that goes directly from the source to the detector? There are at least three possible explanations. First, if from the source one-half of the wave goes in each direction, and if the magnitude of $r_{\mathrm{DP}} \approx 1$, almost all of the energy of the original apically going wave eventually goes basally. Furthermore, if the magnitude of r_{st} is greater than zero, some of the original basally going wave gets delayed due to multiple reflections, then the energy that arrives after the first wave can be greater than the energy of the first wave.

Second, if from the source one half of the wave energy goes in each direction and if there is energy creation apical to the source, i.e., the effective $|r_{\mathrm{DP}}| > 1$, then, independently of the value of r_{st}, more energy will arrive later than earlier.

The third possibility that would allow more energy to arrive late than early assumes that, from the source, more energy is emitted apically than is emitted basally. If the impedance that the source sees in the apical direction is different than the impedance that the source sees in the basal direction, then the assumption that one-half of the energy goes in each direction would obviously be false.

Our data suggests that increasing the a_2/a_1 ratio might increase the energy emitted apically relative to that emitted basally. Both the $f_d(-1)$ and the $f_d(-2)$ show an increase in both delay and the presence of nulls as a_2/a_1 becomes greater than 2 or 3. Since this effect does not depend on the absolute values of the DPs [the $f_d(-2)$ DP is generally 10–15 dB less than the $f_d(-1)$ DP for the same level of primaries], the effect does not seem to be simply due to the nonlinearity of r_{DP}.

However, because it is a common observation that a null can occur for a particular value of the input levels at constant f_2, f_1, f_d, and a_2/a_1 when both levels are varied together, the ratio of a_2/a_1 is not the only factor that determines the energy returned from the DP place. To be able to explain all of the details of the generation of nulls one would have to know the level dependence of the reflection coefficient r_{DP}. The interference pattern seen in stimulus frequency emissions disappears with increasing level, suggesting that the effective reflection coefficient at the emission site decreases as the level increases. Other studies imply that the cochlear reflection coefficient increases with decreasing input level (e.g., Allen *et al.*, 1995). This correlates well with the observation that the measured DP delay regularly increases as the input levels ($a_1 = a_2$) and, therefore, the DP levels decrease (Kimberley *et al.*, 1993).

The features of the nulls that are replicated with this simple model are: (a) The null distribution of broad maxima with narrow minima shows a maximum null density which is approximately the same for the two lower-frequency DPs, $f_d(-1)$ and $f_d(-2)$. (b) The null density is the same whether f_2 is held constant, f_1 is held constant, or f_1/f_2 is held constant (Piskorski *et al.*, 1995). (c) Since, for the two higher-frequency DPs, reflections from the DP place would not reach the ear canal microphone, this model predicts that the fixed time delays for the $f_d(+2)$ and the $f_d(+3)$ DPs would not depend upon a_2/a_1 and these two higher-frequency DPs would not show nulls.

Other features in the phase data that can be explained by the model are (a) It is possible that the later wave can have more energy than the first wave, i.e., it explains the increase in the fixed time delay as a_2/a_1 exceeds a critical value. (b) It explains the commonly observed 2π bifurcations in phase upon changing level (such as shown in Fig. 7 where the phases at the two lowest input levels are different from the phases measured at the higher levels by 2π for $f_d < 2$ kHz). (c) Also, phase patterns like the one in Fig. 5 (left) are exactly what one would observe when a wave with more delay increases in relative amplitude to an early wave (as a function of the change in a_1). (d) Assuming r_{DP} increases with

decreasing amplitude of the DP, the increasing phase delay with decreasing a_1 and a_2 at constant a_2/a_1 ratio is expected.

E. Base versus apex

There seem to be differences in the DP generation mechanics when $f_2 \leqslant 1$ kHz versus when $f_2 \geqslant 2$ kHz. When $f_2 \leqslant 1$ kHz, the two lower-frequency DPs, $f_d(-1)$ and $f_d(-2)$, do not have a distinctive $f_d^{\max}$ that differs from f_2. For $f_2 \geqslant 2$ kHz, $f_d^{\max}$ is defined by a second cochlear map (Allen and Fahey, 1993). The two higher-frequency DPs, $f_d(+2)$ and $f_d(+3)$, behave in the opposite fashion. They do show a distinct $f_d^{\max}$ different from f_2 when $f_2 \leqslant 1$ kHz. When $f_2 \geqslant 2$ kHz, they do not. Some of this effect is probably due to the increasing of the input impedance Z_{in} as the frequency decreases below 1 kHz (Allen, 1986; Rosowski *et al.*, 1986). This means that, for a given value of a_1 in the ear canal, relatively less f_1 energy gets to the nonlinearity as f_1 decreases.

When $f_2 \leqslant 1$ kHz, the $f_d(+2)$ and the $f_d(+3)$ phase varies less with respect to a_1. Since low-frequency neural frequency tuning curves are less sharply tuned, the change in the DP generation properties might be due to the same mechanics that changes the low-frequency neural tuning.

Another observation that may be relevant to the difference of generation mechanics in the base versus apex is that there is broad maximum in the phase of $f_d(+2)$ and $f_d(+3)$ [see Fig. 5 (right panel) and Fig. 8] that is centered about $f_1 \approx 1$ kHz. This is a common feature in the data of the two higher-frequency DPs and it seems to be a function of f_1.

IV. SUMMARY

A. Phase observations that are a property of saturating nonlinearities

Much of the phase data discussed in this study may be explained by the basic properties of saturating nonlinearities. In Fig. 12 the phase of the $f_d(-2)$ DP is about π radians less than the phase of the $f_d(-1)$ DP over the range where phase was measured. This is also true when the phases are plotted versus f_1. In simulations we have found a phase difference of π between these two DPs to be a generic property of the saturating nonlinearities commonly used in modeling auditory phenomena. This was commented upon by Schroeder (1969) for the special case of the $f_d(-1)$ DP phase when he compared the psychophysical phase with that expected from either an expansive nonlinearity or a compressive nonlinearity and found a π phase shift that showed that the source of the $f_d(-1)$ DP is most likely due to a compressive nonlinearity. The deviations from π for successive order DPs would be a measure of the phase difference of f_1 vs f_2 at the site of the nonlinearity. A simple way to understand this is to imagine that the input/output characteristic of a saturating nonlinearity is described in terms of a power series built of odd powers only. The successive terms alternate sign. Each successive term is the dominant term in the next order DP. The change in sign leads to a change in π radians in the phase. For example, the dominant term generating the $f_d(-1)$ DP is the cubic term x^3 and the dominant term generating the $f_d(-2)$ is x^5. In a power series expansion of the input/output characteristic of a saturating nonlinearity the coefficient of x^3 is opposite in sign to the coefficient of x^5.

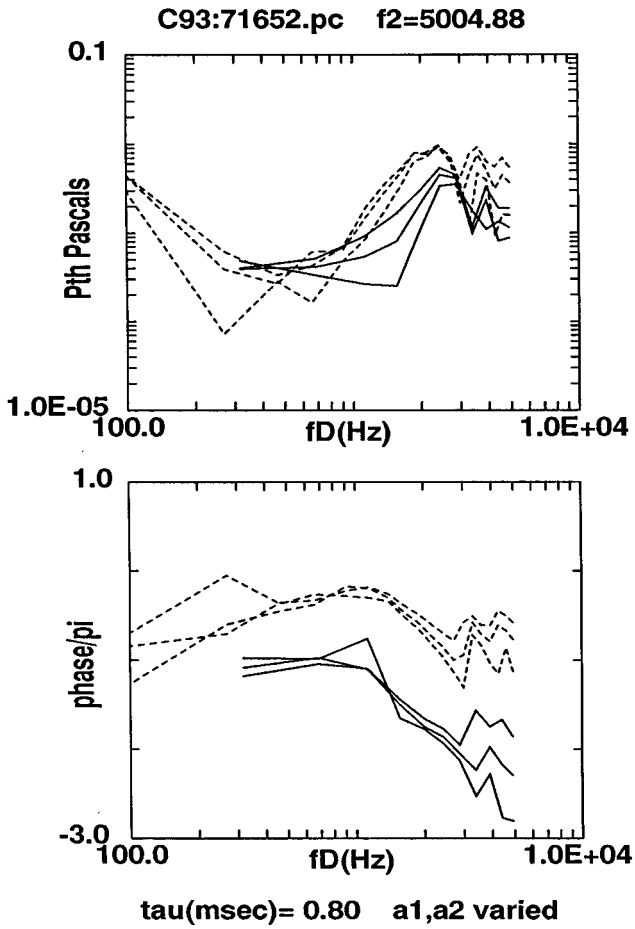


FIG. 12. The three highest magnitude and phase curves (dashed lines) are the $f_d(-1)$ DP and the three lowest (solid lines) are the $f_d(-2)$ DP. The nulls in the 3 to 4 kHz range overlap. These DPs were measured simultaneously (f_1 was decreased by the same amount for each) and, therefore, the point density in the $f_d(-2)$ DP plots is about one-half the point density in the $f_d(-1)$ plots. a_1 and a_2 are decremented together from 1.11 to 0.54 Pa. As the levels decrease the frequency of the nulls decrease and the changing pattern in the magnitude curve for the $f_d(-1)$ DP blends into the pattern for the $f_d(-2)$ DP.

The phase of a DP generated by a simple saturating nonlinearity is independent of input level(s). If it is assumed that the DP observed in the ear canal is the sum of wavelets from an ensemble of distributed saturating nonlinearities along the cochlea, and if it is assumed that the relative amplitudes of the wavelets are independent of level, then the phase of the DP measured in the ear canal would be independent of level. Observation of the magnitude and phase when a_2 is varied and is much less than $a_1 = \text{const}$ seem to fit the above assumption.

B. Phase observations explainable by two delays

For the two higher-frequency DPs [$f_d(+2)$ and $f_d(+3)$] and, under the condition that $a_2/a_1 \leqslant 2$, for the two lower-frequency DPs [$f_d(-1)$ and $f_d(-2)$] the phase has a

level independent delay that is approximately equal to $1/f_2$. It appears that the DPs are generated at the place of the higher-frequency input tone.

However, as a_2/a_1 becomes greater than 2 or 3, the delay that is measured for the two lower-frequency DPs becomes greater. This fact, combined with the observation of nulls in the amplitudes, 2π level-dependent phase bifurcations, and phase patterns such as in Fig. 5 for the two lower-frequency DPs suggest that there is a second ''source'' of the DPs that has a greater delay. We have shown that the nulls can be modeled by assuming that there is reflection of the DP from the f_d place.

ACKNOWLEDGMENTS

The authors would like to thank L. Stover and S. Neely for discussions on the filter function and the IFFT-waveform technique and to thank C. Shera and C. Talmadge for reading and helping to revise a late draft of this work.

Allen, J. B. **(1980)**. ''Cochlear micromechanics—A physical model of transduction,'' J. Acoust. Soc. Am. **68**, 1660–1670.

Allen, J. B. **(1983)**. ''Magnitude and phase frequency response to single tones in the auditory nerve,'' J. Acoust. Soc. Am. **73**, 2071–2092.

Allen, J. B. (**1986**). ''Measurement of eardrum acoustic impedance,'' in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, New York), pp. 44–51.

Allen, J. B., and Fahey, P. F. **(1992)**. ''Using acoustic distortion products to measure the cochlear amplifier gain on the basilar membrane,'' J. Acoust. Soc. Am. **92**, 178–188.

Allen, J. B., and Fahey, P. F. **(1993)**. ''A second cochlear frequency map that correlates distortion product and neural tuning measurements,'' J. Acoust. Soc. Am. **94**, 809–817.

Allen, J. B., Shaw, J., and Kimberley, B. P. (**1995**). ''Characterization of the nonlinear ear canal impedance at low sound levels,'' Abstracts of the Eighteenth Midwinter Research Meeting, Association for Research in Otolaryngology, edited by G. R. Popelka (unpublished).

Bode, H. W. (**1945**). *Network Analysis and Feedback Amplifier Design* (Van Nostrand, New York).

Born, M., and Wolf, E. (**1975**). *Principles of Optics* (Pergamon, New York), 5th ed.

Bowman, D. M., Brown, D. K., Eggermont, J. J., and Kimberley, B. P. **(1997)**. ''The effect of sound intensity in f_1-sweep and f_2-sweep distortion product otoacoustic emissions phase delay estimates in human adults,'' J. Acoust. Soc. Am. **101**, 1550–1559.

Brown, A. M., and Gaskill, S. A. (**1990**). ''Can basilar membrane tuning be inferred from distortion measurement?,'' in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (Springer-Verlag, New York), pp. 164–169.

Brown, A. M., Gaskill, S. A., Carlyon, R. P., and Williams, D. M. **(1993a)**. ''Acoustic distortion as a measure of frequency selectivity: Relation to psychophyical equivalent rectangular bandwidth,'' J. Acoust. Soc. Am. **93**, 3291–3297.

Brown, A. M., Gaskill, S. A., and Williams, D. M. **(1992)**. ''Mechanical filtering of sound in the inner ear,'' Proc. R. Soc. London, Ser. B **250**, 29–34.

Brown, A. M., Harris, F. P., and Beveridge, H. **(1996)**. ''Two sources of acoustic distortion products from the human cochlea,'' J. Acoust. Soc. Am. **100**, 3260–3267.

Brown, A. M., Williams, D. M., and Gaskill, S. A. **(1993b)**. ''The effect of aspirin on cochlear mechanical tuning,'' J. Acoust. Soc. Am. **93**, 3298–3307.

Fahey, P. F., and Allen, J. B. **(1985)**. ''Nonlinear phenomena as observed in the ear canal and at the auditory nerve,'' J. Acoust. Soc. Am. **77**, 599–612.

Fahey, P. F., and Allen, J. B. (**1986**). ''Characterization of cubic intermodulation distortion products in the cat external auditory meatus,'' in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, New York), pp. 314–321.

Fahey, P. F., and Allen, J. B. (**1988**). ''Power law features of acoustic distortion product emissions,'' in *Basic Issues in Hearing*, edited by H. Duifhuis, J. W. Horst, and H. P. Wit (Academic, London), pp. 124–131.

He, N., and Schmiedt, R. A. **(1993)**. ''Fine structure of $2f_1-f_2$ distortion product emissions: changes with primary level,'' J. Acoust. Soc. Am. **94**, 2659–2669.

He, N., and Schmiedt, R. A. (**1996**). ''On the generation site of the fine structure of $2f_1-f_2$ acoustic distortion product in the human ear,'' Abstracts of the Nineteenth Midwinter Research Meeting, Association for Research in Otolaryngology, edited by G. R. Popelka (Association for Research in Otolaryngology, Des Moines, IA).

Hubbard, A. **(1993)**. ''A traveling-wave amplifier model of the cochlea,'' Science **259**, 68–71.

Kim, D. O., and Molnar, C. E. **(1979)**. ''A population study of cochlear nerve fibers: comparison of spatial distributions of average-rate and phase-locking measures of responses to single tone,'' J. Neurophysiol. **42**, 16–30.

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.

Matthews, J. W., and Molnar, C. E. (**1986**). ''Modeling of intracochlear and ear canal distortion product $2f_1-f_2$,'' in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, New York), pp. 258–265.

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.

Piskorski, P., Long, G. R., Talmadge, C. L., and Tubis, A. (**1995**). ''Origin of the fine structure of the distortion product emissions in the human ear,'' Abstracts of the Eighteenth Midwinter Research Meeting, Association for Research in Otolaryngology, edited by G. R. Popelka (Association for Research in Otolaryngology, Des Moines, IA).

Rhode, W. S. (**1980**). ''Cochlear partition vibration-recent views,'' J. Acoust. Soc. Am. **67**, 1696–1703.

Rosowski, J. J., Carney, L. H., Lynch, T. J. III, and Peake, W. T. (**1986**). ''The effectiveness of external and middle ears in coupling acoustic power into the cochlea,'' in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, New York), pp. 3–12.

Ruggero, M. A., and Rich, N. C. **(1991)**. ''Application of a commercially-manufactured Doppler-shift laser velocimeter to the measurement of basilar-membrane vibration,'' Hearing Res. **51**, 215–230.

Ruggero, M. A., Robles, L., and Rich, N. C. **(1992)**. ''Two-tone suppression in the basilar membrane of the cochlea: mechanical basis of auditory-nerve rate suppression,'' J. Neurophysiol. **68**, 1087–1099.

Schroeder, M. R. **(1969)**. ''Relation between critical bands in hearing and the phase characteristics of the cubic difference tones,'' J. Acoust. Soc. Am. **46**, 1488–1492.

Stover, L., Neely, S. T., and Gorga, M. P. **(1996)**. ''Latency and multiple sources of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **99**, 1016–1024.

Sun, X.-M., Schmiedt, R. A., He, N., and Lam, C. F. **(1994)**. ''Modeling the fine structure of the $2f_1-f_2$ acoustic distortion product. I. Model development,'' J. Acoust. Soc. Am. **96**, 2166–2174.

Talmadge, C., Piskorski, P., Tubis, A., and Long, G. (**1996**). ''Evidence for multiple spatial origins of the fine structure of distortion product otoacoustic emissions in humans, and its implications experimental and modeling results,'' Abstracts of the Nineteenth Midwinter Research Meeting, Association for Research in Otolaryngology, edited by G. R. Popelka (Association for Research in Otolaryngology, Des Moines, IA).

Talmadge, C., Tubis, A., Piskorski, P., and Long, G. **(1995)**. ''Modeling distortion product otoacoustic emission fine structure in humans,'' J. Acoust. Soc. Am. **97**, 3413.

Voss, S. E., and Allen, J. B. **(1994)**. ''Measurement of acoustic impedance and reflectance in the human ear canal,'' J. Acoust. Soc. Am. **95**, 372–384.

Whitehead, M. L., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. **(1994)**. ''Measurement of otoacoustic emissions for hearing assessment,'' IEEE Eng. Med. Biol. Mag. **13**, 210–226.

Whitehead, M. L., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. **(1995)**. ''Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. II. Asymmetry in L_1 and L_2 space,'' J. Acoust. Soc. Am. **97**, 2359–2377.

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.

Zweig, G., and Shera, C. A. **(1995)**. ''The origin of periodicity in the spectrum of evoked otoacoustic emissions,'' J. Acoust. Soc. Am. **98**, 2018–2047.

Zwicker, E. (**1981**). ''Cubic difference tone level and phase dependence on frequency and level of primaries,'' *Psychological, Physiological and Behavioral Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft U.P., Delft, The Netherlands), pp. 268–273.

Zwicker, E., and Harris, F. P. **(1990)**. ''Psychoacoustical and ear canal cancellation of $(2f_1-f_2)$-distortion products,'' J. Acoust. Soc. Am. **87**, 2583–2591.

Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers[a)]

Torsten Dau[b)] and Birger Kollmeier
Carl von Ossietzky Universität Oldenburg, Graduiertenkolleg Psychoakustik, AG Medizinische Physik, D-26111 Oldenburg, Germany

Armin Kohlrausch
IPO Center for Research on User-System Interaction, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

(Received 28 June 1996; accepted for publication 4 August 1997)

This paper presents a quantitative model for describing data from modulation-detection and modulation-masking experiments, which extends the model of the ''effective'' signal processing of the auditory system described in Dau *et al.* [J. Acoust. Soc. Am. **99**, 3615–3622 (1996)]. The new element in the present model is a modulation filterbank, which exhibits two domains with different scaling. In the range 0–10 Hz, the modulation filters have a constant bandwidth of 5 Hz. Between 10 Hz and 1000 Hz a logarithmic scaling with a constant Q value of 2 was assumed. To preclude spectral effects in temporal processing, measurements and corresponding simulations were performed with stochastic narrow-band noise carriers at a high center frequency (5 kHz). For conditions in which the modulation rate (f_{mod}) was smaller than half the bandwidth of the carrier (Δf), the model accounts for the low-pass characteristic in the threshold functions [e.g., Viemeister, J. Acoust. Soc. Am. **66**, 1364–1380 (1979)]. In conditions with $f_{\mathrm{mod}} > \Delta f/2$, the model can account for the high-pass characteristic in the threshold function. In a further experiment, a classical masking paradigm for investigating frequency selectivity was adopted and translated to the modulation-frequency domain. Masked thresholds for sinusoidal test modulation in the presence of a competing modulation masker were measured and simulated as a function of the test modulation rate. In all cases, the model describes the experimental data to within a few dB. It is proposed that the typical low-pass characteristic of the temporal modulation transfer function observed with wide-band noise carriers is not due to ''sluggishness'' in the auditory system, but can instead be understood in terms of the interaction between modulation filters and the inherent fluctuations in the carrier. [S0001-4966(97)05611-7]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Mk [JWH]

INTRODUCTION

Temporal resolution in the auditory system, or the ability to resolve dynamic acoustic cues, is very important for the processing of complex sounds. A general psychoacoustical approach to describing temporal resolution is to measure the threshold for detecting changes in the amplitude of a sound as a function of the rate of the changes. The function which relates threshold to modulation rate is called the temporal modulation transfer function (TMTF) (Viemeister, 1979). The TMTF might provide important information about the processing of temporal envelopes. Since the modulation of a sound modifies its spectrum, wide-band noise is often used as a carrier signal in order to prevent subjects from using changes in the overall spectrum as a detection cue; modulation of white noise does not change its long-term spectrum (e.g., Burns and Viemeister, 1981). The subject's sensitivity for detecting sinusoidal amplitude modulation of a broadband noise carrier is high for low modulation rates and decreases at high modulation rates. It is therefore often argued that the auditory system is ''sluggish'' in following fast temporal envelope fluctuations. Since this sensitivity to modulation resembles the transfer function of a simple low-pass filter, the attenuation characteristic is often interpreted as the *low-pass characteristic* of the auditory system. This view is reflected in the structure of a widely accepted model for describing the TMTF (Viemeister, 1979).

It is often argued that the auditory filters play a role in limiting temporal resolution (e.g., Moore and Glasberg, 1986), especially at frequencies below 1 kHz where the bandwidths of the auditory filters are relatively narrow, leading to longer impulse responses or ''ringing'' of the filters. However, the response of auditory filters at high center frequencies is too fast to be a limiting factor in most tasks of temporal resolution (Ronken, 1970; Green, 1973). Thus there must be a process at a more central level of the auditory system than the peripheral auditory filters which limits temporal resolution and causes the ''sluggishness'' in following fast modulations of the stimulus envelope.

Results from several studies concerning modulation masking, however, are not consistent with the idea of only one broad filter, as reflected in the TMTF. Houtgast (1989) designed experiments to estimate the degree of frequency selectivity in the perception of simultaneously presented am-

[a)]Part of this research was presented at the 129th meeting of the Acoustical Society of America [T. Dau, B. Kollmeier and A. Kohlrausch, ''Modeling modulation perception: modulation low-pass filter or modulation filterbank?,'' J. Acoust. Soc. Am. **97**, 3273 (A) (1995)].

[b)]Corresponding author. Electronic mail: torsten@medi.physik.uni-oldenburg.de

plitude modulation, using broadband noise as a carrier. Using narrow bands of noise as the masker modulation, the modulation-detection threshold function showed a peak at the masker modulation frequency. This indicates that masking is most effective when the test modulation frequency falls within the masker-modulation band. In the same vein, Bacon and Grantham (1989) found peaked masking patterns using sinusoidal masker modulation instead of a noise band. Fassel (1994) found similar masking patterns using sinusoids at high frequencies as carriers and sinusoidal masker modulation.

For spectral tone-on-tone masking, effects of frequency selectivity are well established and are associated with the existence of independent frequency channels. When translated to the modulation-frequency domain, the data of Houtgast and of Bacon and Grantham suggest the existence of modulation-frequency specific channels at a more central stage in the auditory system than the peripheral auditory filters. Yost *et al.* (1989) also suggested amplitude modulation channels to explain their data on modulation-detection interference and to account for the formation of auditory ''objects'' based on common modulation. Similarly, Martens (1982) proposed that the auditory system realizes some kind of short-term spectral analysis of the temporal waveform of the signal's envelope.

Modulation-frequency specificity has also been observed in different physiological studies of neural responses to amplitude modulated tones (Creutzfeldt *et al.*, 1980; Langner and Schreiner, 1988; Schreiner and Urbas, 1988; Langner, 1992). Langner and Schreiner (1988) stated that the auditory system contains several levels of systematic topographical organization with respect to the response characteristics that convey temporal modulation aspects of the input signal. A general reduction in the temporal activity patterns of neural elements along the auditory pathway was described as the most basic temporal organizational feature. That is, the temporal resolution of the auditory nerve (Palmer, 1982) appears to be higher than at any other processing level. The highest best modulation frequencies (BMF) found in the inferior colliculus (IC)—which is about 1000 Hz—are still comparable with the temporal resolution of auditory nerve fibers (Schreiner and Langner, 1984; Langner and Schreiner, 1988). However, the majority of units in the IC are tuned to modulation frequencies well below the upper frequency limit given by the auditory nerve. All estimates of temporal resolution in the IC were found to be higher than estimates in the auditory cortex (which are in the range of BMF=0–20 Hz in cats). Thus the auditory cortex seems to be limited in its ability to follow fast temporal changes in the input envelope. On the other hand, the cortex seems to be capable of processing slow modulations like rhythmlike envelope fluctuations (Creutzfeldt *et al.*, 1980). A further organizational level of the temporal processing is reflected by differences found in various subdivisions of auditory nuclei. For example, Langner and Schreiner (1988) found a highly systematically organized *map* of best modulation frequencies within the IC of the cat. Overall, Langner and Schreiner (1988) concluded that temporal aspects of a stimulus, such as envelope variations, represent a major organizational principle of the auditory system, that complements the well-established spectral (tonotopic) and binaural organization.

The present psycho-acoustical study further analyzes the processing of amplitude modulation in the auditory system. The goal is to gather more information about modulation-frequency selectivity and to set up corresponding simulations with an extended version of a model of the ''effective'' signal processing in the auditory system, which was initially developed to describe temporal masking effects (Dau *et al.*, 1996a, b). As already pointed out, in most classical studies of temporal processing, a broadband noise carrier has been applied to determine the TMTF. Unfortunately, the use of broadband noise carriers does not provide direct information about spectral effects in temporal processing: Broadband noise excites a wide region of the basilar membrane, leaving unanswered the question of what spectral region or regions are being used to detect the modulation. For this reason, measurements and corresponding simulations with stochastic narrow-band noises as carriers at a high center frequency were performed, as was done earlier by Fleischer (1982a, 1983). At high center frequencies, the bandwidth of the auditory filters is relatively large so that there is a larger range of modulation rates over which the sidebands resulting from the modulation are not resolved. Instead, the modulation is perceived as a temporal attribute, like fluctuations in loudness (for low modulation rates) or as roughness (for higher modulation rates). The bandwidth of the modulated signal was chosen to be smaller than the bandwidth of the stimulated peripheral filter. This implies that all spectral components are processed together and that temporal effects are dominant over spectral effects.

I. DESCRIPTION OF THE MODEL

In Dau *et al.* (1996a), a model was proposed to describe the effective signal processing in the auditory system. This model allows the prediction of masked thresholds in a variety of simultaneous and nonsimultaneous conditions (Dau *et al.*, 1996b). It combines several stages of preprocessing with a decision device that has the properties of an optimal detector. Since then, the model has also been used to predict speech perception tasks, such as automatic speech recognition and speech quality evaluation (cf. Holube and Kollmeier, 1996; Kollmeier *et al.*, 1996). Figure 1 shows the extended model that is proposed to describe experimental data on modulation perception. Instead of the implementation of the basilar-membrane model developed by Strube (1985), as used in Dau *et al.* (1996a), the gammatone filterbank model of Patterson *et al.* (1987) is used to simulate the bandpass characteristics of the basilar membrane. The gammatone filterbank has the advantages that its algorithm is much more efficient than the Strube model and that the bandwidths more closely match estimates of auditory-filter bandwidths. The signal at the output of a single filter of the gammatone filterbank is half-wave rectified and low-pass filtered at 1 kHz, as in the model described in Dau *et al.* (1996a).

The subsequent nonlinear adaptation stage is a slightly modified version (Münkner, 1993) of the adaptation stage (Püschel, 1988) implemented within the masking model of Dau *et al.* (1996a). In this modified version the amplitude of

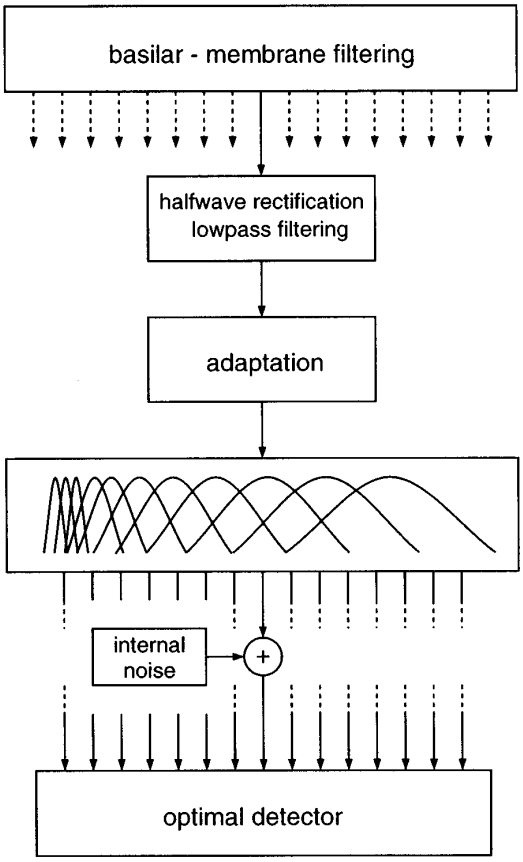


FIG. 1. Block diagram of the psycho-acoustical model for describing modulation-detection data with an optimal detector as decision device. The signals are preprocessed, subjected to adaptation, filtered by a modulation filterbank and finally added to internal noise; this processing transforms the signals into their internal representations.

the onset response was limited to a value of maximally 10 times the value of the steady state response of the stage (for details see Münkner, 1993).[1] With regard to the transformation of envelope variations of the signal, the adaptation stage transforms rapid input variations (as compared with the time constants of the low-pass filters) *linearly*. If these changes are slow enough then, because of the time constants of the model, the gain is also changed. Each element within the adaptation model combines a static compressive nonlinearity with a higher sensitivity for fast temporal variations (for details, see Dau *et al.*, 1996a).

The following stage in the model, as shown in Fig. 1, contains the most substantial changes compared to the model described in Dau *et al.* (1996a). Instead of the low-pass filter with a cutoff frequency of 8 Hz, a linear filterbank is assumed to further analyze the amplitude changes of the envelope. This stage will be called the modulation filterbank throughout this paper. A first implementation of such a modulation filterbank was presented in Fassel and Püschel (1993) and Münker and Püschel (1993). The implementation of this stage is in contrast to the signal processing within other models in the literature (e.g., Viemeister, 1979; Forrest and Green, 1987).

It is postulated within the present model that the modulation filterbank exhibits two domains with different scaling. Figure 2 shows the transfer functions of the modulation filters. In the range 0–10 Hz a constant bandwidth of 5 Hz is assumed. The lowest modulation filter represents a low-pass filter with a cutoff frequency of 2.5 Hz. From 10 Hz up to 1000 Hz a logarithmic scaling with a constant Q value of 2 is assumed.[2] The spacing in the modulation-frequency domain resembles the spacing of critical bands in the audio-frequency domain. Within the model only the (Hilbert) *envelope* of the modulation filter outputs for center frequencies above 10 Hz is further examined, introducing a nonlinearity in the processing of amplitude modulation.[3] For filters with a lower center frequency it is not reasonable to extract the Hilbert envelope from the signal, because the distinction between carrier and envelope becomes ambiguous due to the large relative bandwidth of these filters. Furthermore, the successful description of masking data by the original model version in Dau *et al.* (1996b) suggests that use is made of information about modulation phase at low modulation rates. In this model the signal envelope was analyzed by the simple 8-Hz low-pass filter and this filtering preserves all information about the modulation phase for low modulation frequencies. The present model thus tries to find a ''link'' between the description of phenomena of modulation detection and those of the more common signal detection.

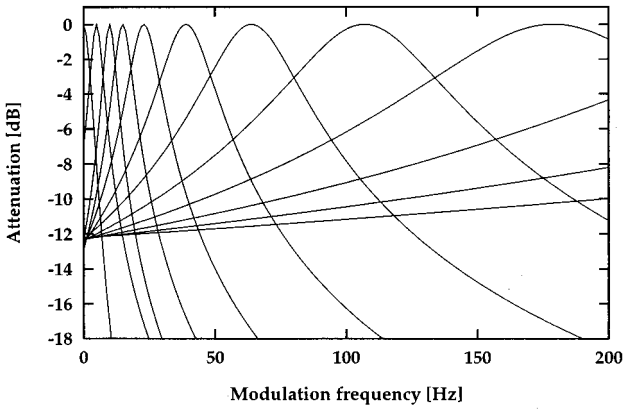


FIG. 2. Transfer functions of the modulation filters. In the range 0–10 Hz the functions have a constant bandwidth of 5 Hz. Between 10 and 1000 Hz a logarithmic scaling with a constant Q value of 2 is applied. Only the range from 0 to 200 Hz is plotted.

The output of the ''preprocessing'' stages can now be interpreted as a three-dimensional, time-varying activity pattern. Limitations of resolution are again simulated by adding internal noise with a constant variance to each modulation filter output.[4] The internal noises at the outputs of the different modulation channels are assumed to be independent of each other. For stochastic input signals, the outputs of the modulation channels are not (fully) uncorrelated because of the overlap of the modulation filters. The transformed signal after the addition of noise is called the internal representation of the signal. The decision device is realized as an optimal detector in the same way as described in Dau *et al.* (1996a, b). There, the decision device of the model was first described for masking conditions using sinusoidal test signals presented in a frozen-noise masker. In each interval of a simulated 3-interval forced-choice (3IFC) adaptive paradigm, the difference between the current representation and the ''stored'' internal representation of the deterministic

masker was calculated, leading to two intervals containing only internal noise and one interval containing the nonlinearly transformed signal plus internal noise. To apply the model to random noise, a different sample of the masker was presented in each interval; the ''reference'' representation was modeled by calculating the mean internal representation of several masker samples (cf. Dau *et al.*, 1996b). This averaged reference was computed once before the adaptive procedure was started. During the adaptive procedure, the three difference representations (in a trial) were affected by the statistical properties of the external noise in addition to the internal noise. Such an algorithm is also used for the present study dealing with stochastic noise as the carrier (and reference) and a sinusoidal modulation (as the test signal). The template is generated in the present study as the normalized difference between an averaged suprathreshold internal representation of several modulated carrier samples and the averaged internal representation of the reference alone (cf. Dau *et al.*, 1996b). The decision criterion within the optimal decision stage (see Green and Swets, 1966) is given by the difference between the largest cross-correlation coefficient of the two carrier-alone representations with the template and the correlation value between the representation in the signal interval and the template. When this difference is smaller than the limit of resolution determined by the internal noise, the test modulation is not detected (for details, see Dau *et al.*, 1996b).[5]

II. METHOD

A. Procedure and subjects

Modulation detection thresholds were measured and simulated using an adaptive 3IFC procedure. The carrier was presented in three consecutive intervals separated by silent intervals of 300 ms. In one randomly chosen interval the carrier was sinusoidally amplitude modulated. In the other intervals it was unmodulated. The subject's task was to specify the interval containing the modulation. During a threshold run, the modulation depth in dB ($20 \log m$), was adjusted using a 2-down 1-up rule (Levitt, 1971) which provides an estimate of the modulation depth necessary for 70.7% correct responses. The step size was 4 dB at the start of a run and was divided by 2 after every two reversals of the modulation depth until the step size reached a minimum of 1 dB, at which time it was fixed. Using this 1-dB step size, 10 reversals were obtained and the median value of the modulation depths at these 10 reversals was used as the threshold value. The subjects received visual feedback after each response. The procedure was repeated four times for each signal configuration and subject. All figures show the median and interquartile ranges based on four single measurements. All five subjects had experience in psycho-acoustic measurements and had clinically normal hearing. They were between 23 and 29 years old and participated voluntarily in the study.

B. Apparatus and stimuli

All acoustic stimuli were digitally generated at a sampling frequency of 30 kHz. The stimuli were transformed to analog signals with the aid of a two-channel 16-bit D/A converter, attenuated, low-pass filtered at 10 kHz, and diotically presented via headphones (HDA 200) in a sound-attenuating booth. Signal generation and presentation were controlled by a SUN Workstation using a signal-processing software package developed at the Drittes Physikalisches Institut in Göttingen.

Several modulation-detection and modulation-masking experiments were performed. In most measurements narrow-band Gaussian noise centered at 5 kHz was used as the carrier. In the masking experiment a sinusoidal carrier at 5 kHz was used. The carrier level was 65 dB SPL in both cases. The specific choice of the parameters for the windowing of the stimuli will be described later in the paper when the corresponding experiments are discussed. In the experiments using a noise carrier, an independent sample of noise was presented in each interval. With one exception described below, the noise stimuli were digitally filtered before modulation by setting the magnitude of the Fourier coefficients to zero outside the desired passband.

An amplitude-modulated noise has sidebands that extend $\pm f_m$ Hz from the edges of the passband of the unmodulated noise, where f_m indicates the modulation frequency. In principle it is possible that the detection of modulation is based on spectral changes in the modulated waveform. The usability of these spectral cues depends on frequency region, owing to the relation of frequency difference limens and center frequency (Wier *et al.*, 1977; Eddins, 1993). One way to avoid these spectral cues is to apply the modulation to wide-band noise before bandpass filtering. In the present study this was done for the largest applied carrier bandwidth, 314 Hz, by setting the magnitude of the Fourier coefficients to zero outside the desired passband. This is the same procedure that was applied by Eddins (1993). Thus in this case, the bandwidth of the stimuli is the same regardless of the presence or absence of modulation. By filtering after amplitude modulation, the sidebands introduced by modulation are effectively reduced. The filtering after modulation causes a partially filling in the valleys of the temporal waveform (e.g., Eddins, 1993). However, this technique ensures that spectral cues were not available and the task was purely temporal in nature. In contrast, for the carrier bandwidths of 3 and 31 Hz, no filtering after modulation was applied.

When generating amplitude-modulated narrow-band stimuli, the average power of the modulated signal is increased by $1 + m^2/2$ compared with the unmodulated signal. For large modulation depths, detection might therefore be based on changes in overall intensity rather than on the presence or absence of modulation. To eliminate level cues, the digital waveforms were adjusted to have equal power in each interval of the forced-choice trial.

In most cases sinusoidal test modulation with zero onset phase was applied. In one experiment a complex modulator was used, consisting of five adjacent components of a harmonic tone complex. In each case the carrier and the applied modulators were windowed with a length depending on the particular experiment.

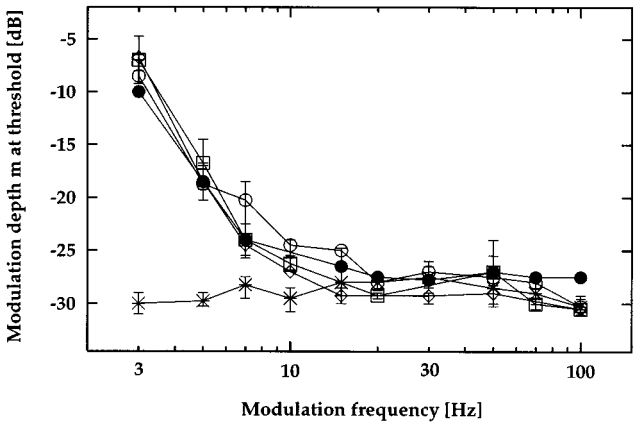


FIG. 3. Modulation-detection thresholds of sinusoidal amplitude modulation as a function of the modulation frequency. The carrier was a 3-Hz-wide running noise at a center frequency of 5 kHz. Carrier and modulation duration: 1 s. Level: 65 dB SPL. Subjects: JV (□); AS (◇); TD (○); optimal detector (●). In addition, the modulation detection thresholds of one subject (TD) for a 5-kHz sinusoidal carrier are indicated by (⋆).

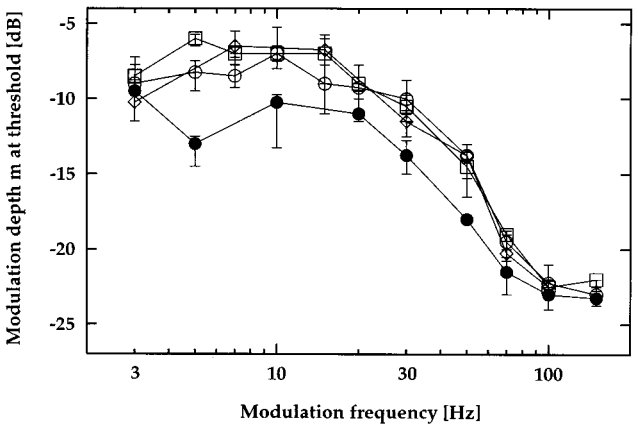


FIG. 4. Modulation-detection thresholds of sinusoidal amplitude modulation as a function of the modulation frequency. The carrier was a 31-Hz-wide running noise at a center frequency of 5 kHz. Carrier and modulation duration: 1 s. Level: 65 dB SPL. Subjects: AS (◇); TD (○); JV (□); optimal detector (●).

III. RESULTS

A. Amplitude-modulation thresholds of narrow-band noise as a function of the carrier bandwidth

In this experiment, TMTFs were measured and simulated for narrow-band noise carriers of bandwidths 3, 31, and 314 Hz, centered at 5 kHz in each condition (cf. Fleischer, 1982a, 1983). Fleischer's experiments were replicated in this study and compared with corresponding simulations carried out with the present model. In contrast to Fleischer, an adaptive threshold procedure was used and the carrier level was somewhat lower (65 dB SPL). For the three noise bandwidths, the corresponding spectrum levels were about 60, 50, and 40 dB SPL. The carrier and the applied sinusoidal modulation had a duration of 1 s. Both were windowed with 200-ms cosine-squared ramps. Figure 3 shows the present experimental results for amplitude modulation detection employing a carrier bandwidth of 3 Hz at a center frequency of 5 kHz. The figure shows the data of three subjects (open symbols) together with the model predictions (closed symbols). For comparison, data obtained by one subject using a sinusoidal carrier at 5 kHz are shown as asterisks. For the sinusoidal carrier, the same stimulus parameters were used as for the noise carrier conditions. The ordinate indicates modulation depth at threshold, and the abscissa represents the modulation frequency. A comparatively high detection threshold is observed at a modulation rate of 3 Hz. This is due to the *inherent* statistical fluctuations of the narrow-band 3-Hz-wide carrier. These inherent fluctuations of the carrier envelope mask the additional periodic 3-Hz test modulation. With increasing modulation frequency, thresholds decrease and converge with those obtained using a sinusoidal carrier at a modulation frequency of 20 Hz. The threshold remains flat up to a modulation frequency of 100 Hz. This finding indicates that the auditory system does not seem too slow or sluggish to follow fast fluctuations in this range. There is very good agreement between the measurements and simulations. The flat threshold function up to 100 Hz contrasts with the conclusions derived from modulation detection data for broadband noise carriers (e.g., Viemeister, 1979) and also contrasts with data in Zwicker (1952) obtained with a sinusoidal carrier. Zwicker found an increase in threshold between 4 and 64 Hz of about 9 dB at a carrier frequency of 4 kHz. However, the present data are in good agreement with more recent data by Fleischer (1982a, 1983), Fassel (1994), Fassel and Kohlrausch (1995), Dau (1996), and Fassel *et al.* (1997), who measured TMTFs with sinusoidal carriers at 5 and 10 kHz.

Flat thresholds up to a modulation frequency of 128 Hz were also observed by Strickland and Viemeister (1997) in an experiment where subjects had to discriminate between AM and quasi-frequency modulation (QFM) applied to a sinusoidal carrier of 4 kHz. Based on additional data on QFM detection, these authors argued that the flatness in their TMTF between 64 and 128 Hz may have been caused by the increasing role of spectral cues and thus did not reflect true temporal processing. Since the assumptions about available cues are of relevance for the interpretation of our data, we will return to the arguments put forward by Strickland and Viemeister in the discussion (Sec. V) of the present paper.

Figure 4 shows thresholds using a narrow-band carrier with a bandwidth of 31 Hz. Again, the modulation depth, m, at threshold was measured and simulated as a function of the test-modulation frequency. The open symbols represent the measured data of three subjects and the filled symbols indicate the simulated thresholds. The threshold at a very low modulation rate (3 Hz) is several dB lower than in the case of the 3-Hz-wide carrier. This decrease is due to the decreasing spectral energy density in the modulation spectrum with increasing bandwidth of the carrier (see the Appendix). In terms of the model, less ''noise energy'' falls into the low-frequency modulation filter which is tuned to the test-modulation frequency. For modulation frequencies larger than half the bandwidth of the noise ($f_{\mathrm{mod}} > \Delta f/2$) thresholds begin to decrease, both in the measurements and in the simulations, so that a high-pass characteristic in the threshold function becomes apparent. However, thresholds decrease more slowly with increasing modulation frequency than the

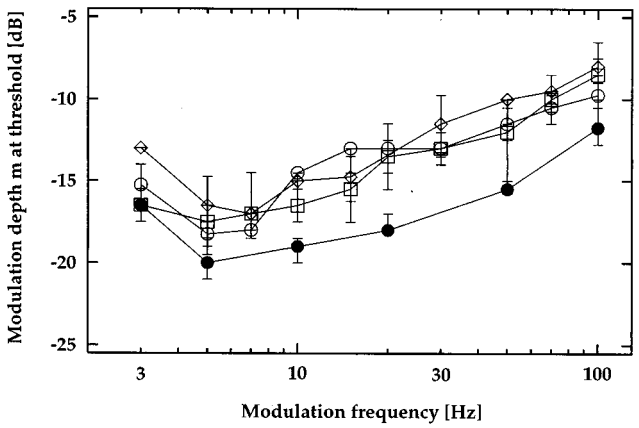


FIG. 5. Modulation-detection thresholds of sinusoidal amplitude modulation as a function of the modulation frequency. The carrier was a 314-Hz-wide running noise at a center frequency of 5 kHz. Carrier and modulation duration: 1 s. Level: 65 dB SPL. Subjects: JV (□); TD (○); AS (◇); optimal detector (●).

spectrum of the inherent envelope fluctuations itself (e.g., Lawson and Uhlenbeck, 1950). The idealized modulation spectrum of a rectangular shaped bandpass noise has a triangular shape which stretches from 0 to Δf Hz (see the Appendix and Lawson and Uhlenbeck, 1950). If the auditory system would be sharply tuned in frequency selectivity for modulation, thresholds would decrease with increasing modulation frequency more or less in parallel with the spectral shape of the modulation spectrum of the carrier (assuming a constant signal-to-noise ratio at the output). Apparently, this is not the case. Hence, even at high modulation rates of 100 and 150 Hz, thresholds have not yet converged with those for the 3-Hz-wide carrier nor with those for the sinusoidal carrier, but are about 5 dB higher. This implies that the relatively slow inherent fluctuations of the 31-Hz-wide carrier make it difficult to detect the higher-frequency test modulation. This phenomenon was also observed by Fleischer (1982a) who referred to it as ''cross-talk'' of the inherent fluctuations of the noise on the added modulation. This effect decreases with increasing rate of the test modulation.

This experiment reveals much about the auditory system's selectivity for modulation frequency. In the model it was necessary to use wide modulation filters ($Q=2$) at high modulation frequencies so that some energy from the low-frequency fluctuations of the ''masker'' leaks through a modulation filter that is tuned to a high modulation frequency (like 150 Hz). This leakage decreases the signal-to-noise ratio and therefore leads to a higher detection threshold at $f_{\mathrm{mod}}=150$ Hz than would be the case for a more sharply tuned filter. Again, there is good agreement between the form of the simulated and measured data.

Figure 5 shows results for the carrier bandwidth of 314 Hz. Thresholds are higher for a modulation rate of 3 Hz than for a rate of 5 Hz. This is probably caused by the use of a gated carrier. Such an effect has been observed in several psycho-acoustical studies (e.g., Viemeister, 1979; Sheft and Yost, 1990). Based on their results it can be assumed that the threshold at 3 Hz would decrease if a continuous carrier had been used instead of a gated one. For modulation frequencies above 7 Hz, thresholds increase by about 3 dB per doubling of the modulation frequency. This threshold pattern agrees well with comparable experimental data of Eddins (1993) for a carrier bandwidth of 400 Hz. The form of the TMTF is similar to the pattern found in ''classical'' measurements of the TMTF using a broadband noise as a carrier, but it has a much lower cutoff frequency (Viemeister, 1979). Overall, the threshold curve is very different from those obtained with smaller carrier bandwidths since the detectability of the test modulation decreases with increasing modulation frequency. Consistent with the data, the simulations also show increasing thresholds with increasing modulation frequency.

For a carrier bandwidth of 314 Hz, the spectrum of the intrinsic fluctuations is relatively flat over the whole range of the test-modulation frequencies. The additional test component falls in the passband of mainly one modulation filter. Assuming a constant decision criterion at threshold, the logarithmic scaling of the modulation filters with center frequencies above 10 Hz leads to an approximately 3-dB increase of modulation depth, m, at threshold per doubling of modulation frequency. In other words, to get the same signal-to-noise ratio at threshold, a greater modulation depth is required with increasing modulation frequency. Thus the apparent modulation low-pass behavior in the model data in Fig. 5 is not explained by assuming a general low-pass characteristic in the auditory system, but is caused by the constant relative width (or logarithmic scaling) of the modulation filters.

Figure 6 gives an illustration of how the signals are internally represented in the model. It shows how the template is derived from the internal representation of suprathreshold test modulation and that of the unmodulated carrier alone. The upper panel shows the three-dimensional internal representation of a 3-Hz-wide carrier alone (centered at 5 kHz). It represents the internal activity as a function of time and center frequency of the modulation filters. The ordinate is scaled in model units (MU). The modulation center frequencies range from 0 to 1000 Hz. Since the total energy within the modulation spectrum of the signal is concentrated at very low modulation rates, only the lowest modulation filters are excited by the input signal. This is indicated by the hatched lines in the figure. At the beginning of the carrier, all modulation filters show a short period of high excitation. This response reflects the step response of the filters to the envelope onset. The middle panel of Fig. 6 shows the internal representation of the carrier, this time sinusoidally modulated with a test-modulation rate of 20 Hz at a highly detectable modulation depth. The test modulation mainly activates the modulation filter tuned to 20 Hz but also stimulates adjacent modulation filters, because of the relatively low modulation-frequency selectivity assumed in the model. Again, the inherent fluctuations of the carrier itself primarily activate the region at low modulation frequencies. However, because of the large spectral separation between the test modulation and the inherent fluctuations of the carrier, there is no interaction between the two components; that is, no competing ''noise'' energy leaks into the transfer range of the test-modulation filter. The lower panel in Fig. 6 gives the

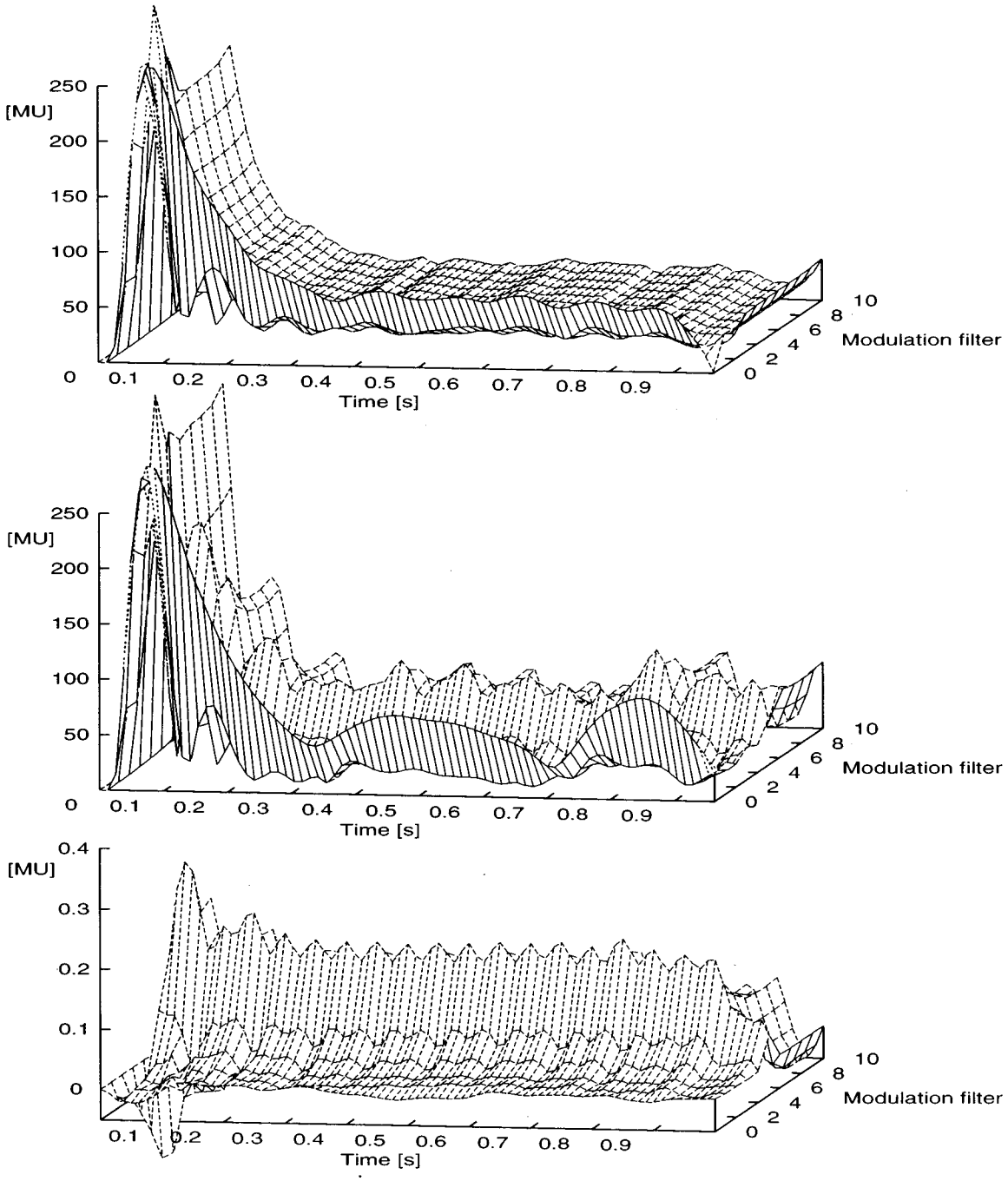


FIG. 6. Generation of the template representation (at the bottom) of a 20-Hz test modulation which was impressed on a 3-Hz-wide running noise carrier centered at 5 kHz. The template is the normalized difference between the mean representation of the carrier plus the suprathreshold modulation (in the middle) and the mean representation of the carrier alone (at the top). The ordinate is scaled in model units (MU).

template which is derived by subtracting the upper panel from the middle one and normalizing the result. As a consequence of the marked separation in modulation frequency, the internal representation of the template contains a representation of the temporal course of the test modulation without interference from the carrier modulation.

B. Amplitude modulation thresholds of third-octave-wide noisebands as a function of the center frequency

In the previous section it was observed that the detection threshold for amplitude modulation depends on the spectral density of the inherent fluctuations of the carrier, when the total energy of the modulated signal is constant. This is examined further in the following experiment using a third-octave-wave noiseband as the carrier. The detection threshold for 25-Hz modulation was measured and simulated as a function of the center frequency of the band. Stimulus parameters were the same as in the experiments of the previous section. In the model, only the output of the peripheral filter centered on the bandpass noise was analyzed. It was further assumed that the scaling of the modulation filters does not change with the peripheral frequency region. Figure 7 shows the modulation depth at threshold as a function of the center

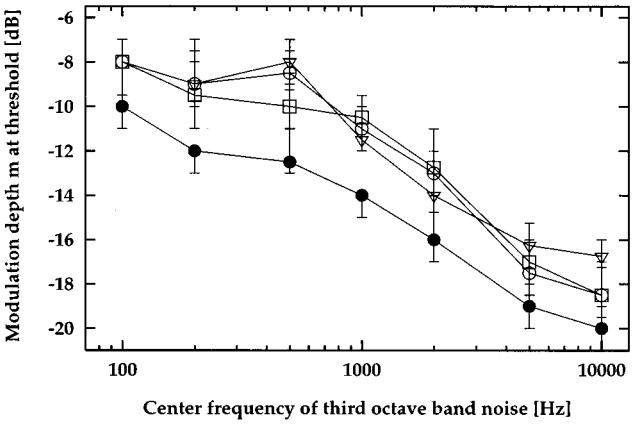


FIG. 7. Modulation-detection thresholds for a 25-Hz modulation as a function of the center frequency of a third-octave-wide noise carrier. Carrier and modulation duration: 1 s. Level: 65 dB SPL. Subjects: Data from Fleischer (1981) (□); TD (○); JV (▽), optimal detector (●).

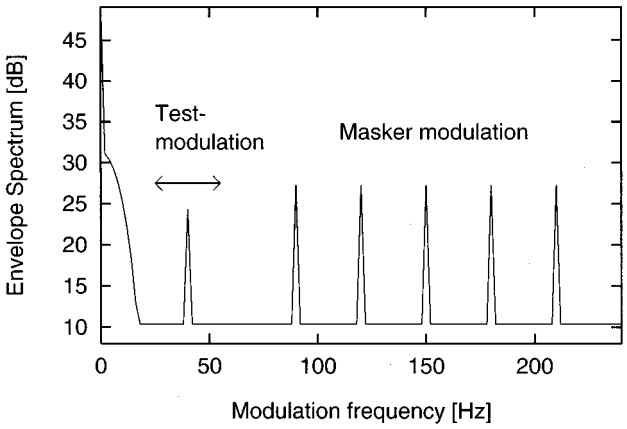


FIG. 8. Logarithmic spectrum of the Hilbert envelope of the stimuli presented in the signal interval of the modulation masking experiment. The signal interval contains the five components of the masking tone complex and the signal component. The subject's task was to detect the signal component.

frequency of the third-octave-wide noiseband. Modulation thresholds decrease with increasing center frequency. The increasing absolute bandwidth results in a decreasing density of inherent low-frequency envelope fluctuations, if the total energy of the modulated stimulus is kept constant. As a consequence, less energy from the random envelope fluctuations of the carrier falls within the passband of the modulation filter tuned to the test-modulation frequency. This leads to decreasing thresholds with increasing center frequency in the model. Apart from the systematic 2- to 3-dB difference in the absolute sensitivity, there is good agreement between the simulated and the measured data.

C. Modulation masking: a harmonic tone-complex masker

In a further experiment concerning modulation-frequency selectivity, a masking paradigm for investigating frequency selectivity in the audio-frequency domain was adopted. It served as a test for spectral analysis in the modulation domain, as opposed to a periodicity analysis. The carrier was a 5-kHz sinusoid. A narrow-band tone complex was used as masker modulation. This complex consisted of the third through seventh components of a harmonic tone complex with a fundamental frequency of 30 Hz, with frequencies of 90, 120, 150, 180, and 210 Hz. The amplitude of each component was 0.16, a value sufficiently low to avoid overmodulation when the test modulation was combined with the tone-complex modulation. In each interval, the starting phase of each spectral component was randomly chosen from a uniform distribution in the range 0–360°. With this choice, the modulating tone complex had a noiselike, but periodic, waveform. A sinusoidal test modulation was imposed on the same carrier. The test modulation was chosen from the range 20–120 Hz. Thus the bandwidth of the modulated signal remained within the bandwidth of the auditory filter centered at 5 kHz. Figure 8 shows schematically the spectral distribution of the masker and test components in the modulation spectrum. The modulated stimuli were presented at a level of 65 dB SPL, and had a duration of 400 ms. Test and masker modulation were present for the whole duration of the carrier and were gated with 20-ms cosine-squared ramps.

The amount of modulation masking as a function of the test-modulation frequency is shown in Fig. 9. The unmasked modulation thresholds, i.e., the thresholds for sinusoidal test modulation without any interfering masker modulation, were used as a reference to evaluate the effect of the modulated masker. These reference thresholds were similar across subjects and were similar to those described in the first experiment of this paper (see Fig. 3), remaining more or less flat up to a modulation frequency of 120 Hz. The ''masking pattern'' was derived by subtracting the unmasked threshold from the masked threshold at each test-modulation frequency. As can be seen from Fig. 9, the amount of masking increases with increasing test-modulation frequency. The difference between the highest and the lowest threshold was more than 10 dB. Note that there is no peak at 30 Hz, the ''missing fundamental.'' This indicates that the masking effect is not determined by the period of the masker modulation. Also, no pronounced peak in threshold is observed for

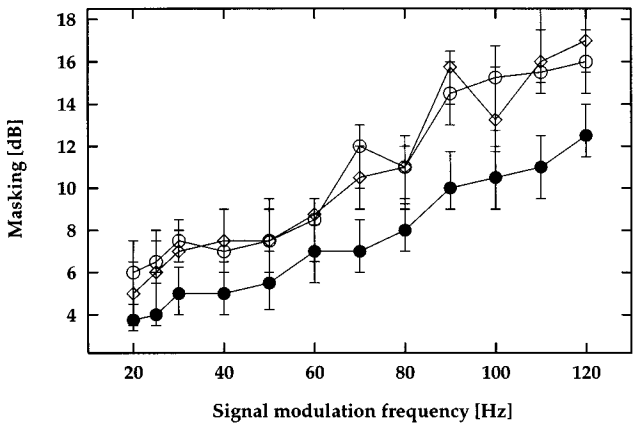


FIG. 9. Amount of modulation masking as a function of the modulation frequency. Carrier: 5-kHz sinusoid, modulation masker: 3rd–7th components of a harmonic tone complex with fundamental $f_0 = 30$ Hz. Level: 65 dB SPL. Subjects: TD (○); JV (◇); optimal detector (●).

modulation frequencies of 90 and 120 Hz, corresponding to the lowest harmonic components of the masker complex. This indicates that no sharp tuning in modulation frequencies occurs at these comparatively high modulation frequencies. An increasing masking effect with increasing test modulation frequency is also seen in the simulations (filled symbols in Fig. 9). If masking effects in the modulation-frequency domain were determined by the periodicity of the stimuli one would expect an increased threshold at the fundamental frequency of the tone complex and at higher harmonics. If, however, the masked threshold of the test signal were mainly determined by the auditory system's frequency selectivity for modulation, one could conclude that the system performs a modulation analysis which is comparable and analogous to the ''critical-band'' filtering on the basilar membrane. The experimental data clearly suggest the latter case. The simulations show good agreement with the experimental data. However, there is a systematic difference of 2–4 dB between the measured and simulated masking patterns. The masked threshold is directly related to the amount of masker energy falling within the passband of the modulation filter tuned to the actual test modulation. For the lowest test modulation rate (20 Hz) there is only a very small masking effect in the model since the modulation filters in the low modulation-frequency region are assumed to be relatively sharply tuned (see Fig. 2). With increasing test-modulation frequency, more and more components of the masker contribute to masking. Also, in the simulations the difference between the highest and the lowest masked threshold amounts to nearly 10 dB. These results further support the notion of modulation-frequency selectivity, although this selectivity seems to be relatively broadly tuned.

IV. COMPARISON WITH PREDICTIONS OF VIEMEISTER'S MODEL FOR MODULATION DETECTION

The modulation filterbank concept differs considerably in its structure from the ''classical'' modulation low-pass filter approach (e.g., Viemeister, 1979). In this section, predictions of the modulation low-pass filter approach are investigated and compared with the performance of the modulation filterbank model.

The structure of Viemeister's model incorporates a predetection bandpass filter (with a bandwidth of $\Delta f = 2000$ Hz) which is followed by a nonlinearity (half-wave rectification) and a low-pass filter. Viemeister fitted the cutoff frequency of the low-pass filter to the TMTF data obtained with a broadband noise carrier. The resulting cutoff frequency was 64 Hz. As a decision variable he suggested the ac-coupled root-mean-square (rms) value of the output of the low-pass filter which was calculated over the duration of the observation interval. The thresholds were defined as the modulation depth necessary to produce a certain average increment (in dB) in the rms value, compared to that for an unmodulated noise.

Figure 10 shows simulated TMTFs for noise carriers of 3-, 31-, 314-, 2000-, and 6000-Hz bandwidth on the basis of Viemeister's model. The narrow-band stimuli were the same as in Figs. 3–5. All curves show a low-pass characteristic

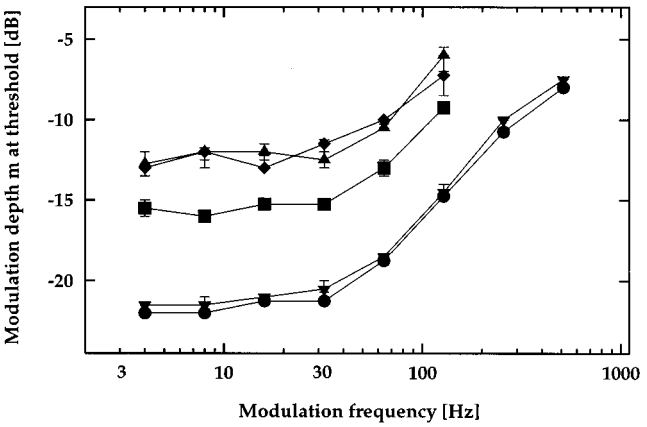


FIG. 10. Simulations on the basis of Viemeister's model. Predicted modulation detection thresholds are shown for five different bandwidths of the noise carrier. Center frequency of the carrier: 5 kHz. Carrier bandwidth: ♦: 3 Hz; ▲: 31 Hz; ■: 314 Hz; ▼: 2000 Hz; ●: 6000 Hz.

with a similar cutoff frequency. This characteristic reflects the influence of the low-pass filter stage. With decreasing carrier bandwidth, the simulated TMTFs shift toward higher thresholds. For the output of Viemeister's model, this increase will be seen for carrier bandwidths that are less than the bandwidth of the predetection filter and greater than the cutoff frequency of the low-pass filter. For these conditions, the ac-coupled rms value of the unmodulated noise carrier at the output of the modulation low-pass stage increases with decreasing carrier bandwidth.

At very low modulation frequencies, the increase in threshold agrees qualitatively with the experimental data. At higher modulation frequencies, however, the model predicts a totally different threshold pattern than that observed experimentally.

While the pattern of the experimental data varies systematically with increasing carrier bandwidth, the model always predicts a low-pass characteristic in the threshold function independent of the carrier bandwidth. Note that the model proposed here provides a better description of the experimental data (cf. see Sec. III A, Figs. 3–5).

Figure 11 shows model predictions of amplitude modu-

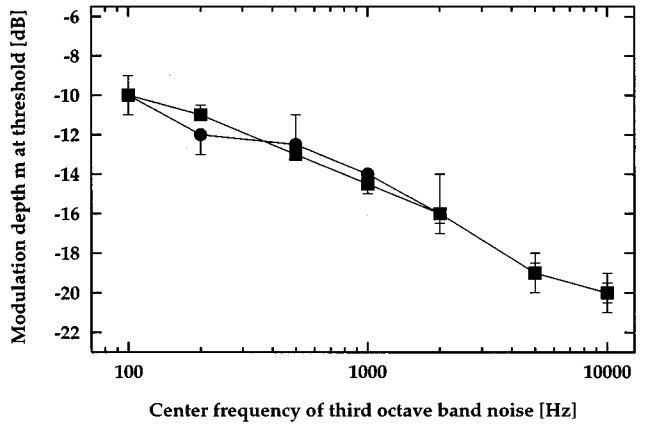


FIG. 11. Simulated modulation detection thresholds for 25-Hz amplitude modulation as a function of the center frequency of the third-octave-wide noise carrier. Viemeister model: (■); modulation filterbank model (●).

lation detection thresholds of third-octave-noise bands as a function of the center frequency. The stimuli were the same as in Sec. III B. The filled squares represent thresholds on the basis of the Viemeister model. Thresholds decrease monotonically with increasing center frequency. This is again caused by the decreasing ac-coupled rms value of the unmodulated noise with increasing center frequency (and increasing linear bandwidth) at the output of the modulation low-pass stage, in the same way as described above. The filled circles in the figure show the simulated data obtained with the modulation filterbank model (replotted from Fig. 7). There is virtually no difference between the predictions of the two models. The predicted threshold functions account well for the data (see Fig. 7).

Finally, model predictions were calculated for the modulation-masking experiment with the harmonic tone complex as the masker. The stimuli were the same as in Sec. III C. The model of Viemeister predicts about 5-dB masking for *all* test-modulation frequencies. This frequency-independent masking is caused by the specific model structure: Because there is only one modulation-frequency specific channel (the output of the low-pass filter) the effect of the masker modulation is the same for all test modulation frequencies at least in combination with the decision algorithm proposed by Viemeister (1979). Hence the experimentally observed increase in modulation masking with decreasing spectral distance between modulation masker and test modulation cannot be described properly with Viemeister's model.

While this paper was being written, a recent article by Strickland and Viemeister (1996) showed that by replacing the rms criterion with a max/min statistic, a single-channel envelope detector can capture some aspects of modulation masking data, so that the effect of masker modulation is not the same at all test modulation rates. However, their model predictions showed a much too sharp ''tuning'' to modulation frequency compared to the relatively broadly tuned masking patterns found in the data (Houtgast, 1989; Bacon and Grantham, 1989). In order to better understand the properties of such statistics, we repeated the simulations shown in Fig. 10 using a max/min decision device. Basically, the change in detector criterion from rms to max/min does not change the main aspects of the previously discussed curves: Independent of carrier bandwidth, all TMTFs have the same shape and increase with increasing modulation frequency. For a reduction of the carrier bandwidth between 2000 and about 30 Hz, the TMTFs are shifted toward higher threshold values. In addition, as already mentioned by Forrest and Green (1987), the max/min statistic is less stable than the rms statistic.

In summary, the analysis of various models proposed in the literature and the simulations from the present study provide a strong indication for a modulation-frequency specific analysis in the auditory system. The modulation filterbank model, which is able to reproduce at least the trend in the data, is one possible realization for this analysis.

V. DISCUSSION

The main goal of this study was to develop a model which describes the effective processing of envelope fluctuations in the auditory system. Experiments concerning modulation detection and modulation masking were performed which suggest that the auditory system realizes some kind of spectral decomposition of the temporal envelope of the signals. There seem to be channels in the auditory system which are tuned to modulation frequency, much like there are channels (critical bands or auditory filters) tuned to spectral frequency.

With regard to the experiments performed and the structure of the model that is inferred from these data, the following points should be discussed: (a) the assumption that the use of spectral cues, effects of peripheral filtering, and off-frequency listening can be neglected for the conditions tested in this paper; (b) the concept of a modulation filterbank as opposed to a modulation low-pass filter in each critical band; and (c) the envelope statistics of the different noise maskers employed and their influence on the thresholds obtained here.

A. Role of spectral cues, peripheral filtering, and off-frequency listening

The experiments in this study have been designed so as to minimize effects of spectral cues and peripheral filtering. The carrier frequency was very high and therefore the bandwidth of the modulated signals was always smaller than the bandwidth of the stimulated peripheral filter. We also argued that at a carrier frequency of 5 kHz, temporal cues are dominant over spectral cues in modulation detection for modulation frequencies up to at least 100 Hz, a view we find supported by measurements with sinusoidal carriers (see Sec. III A, and Fleischer, 1982a, 1983; Fassel and Kohlrausch, 1995, 1996; Dau, 1996; Fassel *et al.*, 1997).

A somewhat different view about the flatness of TMTFs for tonal carriers at 4 kHz was put forward in a recent paper by Strickland and Viemeister (1997). Based on data for discrimination between AM and QFM, and on data for detecting QFM, they argued that the thresholds for AM vs QFM at a modulation frequency of 128 Hz were not just caused by temporal cues (as we assume), but that other, probably spectral cues, were also involved. It is further implied that this may also be the reason for the flatness in tonal TMTFs at 3 and 5 kHz in Fassel and Kohlrausch (1996) and Dau (1996), which would undermine one of the assumptions used in our interpretation. In the following we argue why we find this implication not convincing.

According to Strickland and Viemeister, the sensitivity to temporal cues alone would lead to an increase in thresholds for AM vs QFM discrimination above 64-Hz modulation frequency. Only due to the availability of additional cues, thresholds appear to be flat up to 128 Hz. The usability of these additional cues is derived from experiments measuring QFM detection. At a 4-kHz carrier frequency and 128-Hz modulation frequency, QFM detection thresholds are about 8 dB higher than the thresholds for discriminating AM from QFM. These differences in level make it, in our view, very difficult to see room for a reasonable contribution of nontem-

poral cues, unless it was shown that the psychometric function for QFM detection was sufficiently shallow.

If we compare AM detection thresholds at a wider range of carrier frequencies, the interpretation by Strickland and Viemeister leads to the following view. Data by Zwicker (1952) show that QFM detection thresholds as a function of modulation frequency decrease earlier for lower than for higher carrier frequencies and that the shift in shape is about 1 oct in modulation frequency per 1 oct in carrier frequency (cf. Figs. 10 and 11 in Zwicker, 1952). If we make the parsimonous assumption that the contribution of nontemporal cues to AM vs QFM discrimination follows the same basic rules across carrier frequencies, the modulation frequency at which nontemporal cues start to contribute should be lower than 128 Hz for carrier frequencies below 5 kHz and higher than 128 Hz for carriers above 5 kHz. Our own results on AM detection for sinusoidal carriers do not show such behavior (Fassel and Kohlrausch, 1995; Dau, 1996; Fassel *et al.*, 1997). For all carrier frequencies between 3 and 10 kHz, AM detection thresholds remain flat up to the same modulation frequency of about 100 to 130 Hz.

Another problem in interpreting AM detection thresholds is that observers may increase the relative modulation depth in the AM stimulus by positioning their ''internal observation filter'' away from the carrier frequency in such a way as to better equate the amplitude of the carrier and one of the sidebands (e.g., Goldstein, 1967). The increase in relative modulation depth resulting from listening off frequency would improve performance. We think that for 5-kHz carriers and modulation frequencies in the range 0–100 Hz, it is unlikely that this type of off-frequency listening would be advantageous, and that subjects most likely monitor the internal filter at the carrier frequency.

First of all, the (relative) frequency difference between carrier and one of the sidebands is no larger than 2%, which corresponds to about 0.15 ERB or 0.1 Bark. In order to increase the modulation depth m at the output of an off-frequency filter by 2 dB, this filter would need to have a slope of 6 dB per 100 Hz (assuming a constant slope over the spectrum of the AM stimulus). This corresponds at 5 kHz to slopes of about 33 dB/ERB or 56 dB/Bark, values clearly higher than typical estimates of auditory filter slopes.

In addition, if this effect contributes to AM detection, it should be even stronger in single-sideband detection. If only one sideband and the carrier are available, optimal placing of a filter away from the carrier will increase the degree of modulation even more than is possible for modulation detection, where the (relative) increase of one sideband is accompanied by a decrease of the other sideband. Both for 5-kHz (Dau, 1996) and 10-kHz carriers (Fassel *et al.*, 1997), we found basically flat thresholds for detecting the lower or the upper sideband up to about 100 Hz. Even more important is the observation that sideband detection thresholds for larger spectral distances first increased, before they finally decreased. We take this as an indication that monitoring an off-frequency filter cannot significantly influence TMTFs at 5 kHz for modulation frequencies up to at least 100 Hz.

Alternatively, off-frequency listening could influence modulation detection for narrow-band carriers by the increased internal modulation depth in the region of upward spread of excitation. Such a mechanism has been proposed in the past as one of the sources for the level dependence of AM detection thresholds for sinusoidal carriers (e.g., Zwicker, 1956; Maiwald, 1967b). According to Strickland and Viemeister (1997), it also affects thresholds for low modulation frequencies in the case of bandlimited noise carriers. This conclusion was based on the fact that by adding unmodulated notched noise designed to mask the region of upward spread of a bandlimited noise carrier, modulation-detection thresholds increased by up to 7 dB. Based on this result, it was argued that measuring modulation thresholds without a notched noise would not reveal true temporal processing within the auditory filter centered on the carrier.

Interestingly, the usability of nonlinear upward spread in modulation detection for noise carriers has been addressed theoretically and by model simulations earlier by Maiwald (1967b) and we will recall the relevant points here. In the region of upward spread, both the inherent fluctuations of the noise carrier *and* the applied AM will be enhanced in a similar way. As long as the intrinsic fluctuations of the carrier are the limiting factor for modulation detection, the upward spread region does not allow a better detectability than the on-frequency region. This contrasts to the situation for sinusoidal carriers, where only the applied modulation, but not the limiting (internal) noise are enhanced in the region of upward spread.

Second, if nonlinear upward spread indeed plays such a significant role as stated by Strickland and Viemeister (1997), modulation thresholds for bandlimited noise carriers should strongly increase with decreasing carrier level, since the availability of nonlinear upward spread is strongly reduced at low and medium carrier levels. Data by Maiwald (1967b) for a 127-Hz-wide noise carrier at 1 kHz show that detection thresholds for 4-Hz modulation vary by no more than 2 dB for a level variation of 60 dB. In contrast, the same level variation for a sinusoidal carrier reveals a 15-dB effect. This suggests that the results of Strickland and Viemeister were not primarily due to the masking of the upward spread of excitation.

Following from these considerations we conclude that for the conditions investigated in the present study, monitoring off-frequency filters does not contribute significantly to modulation detection and we can indeed attribute the thresholds to being based on temporal, rather than on spectral cues.

Of course, in modulation-detection conditions with carrier bandwidths larger than a critical band, the influence of peripheral filtering on the processing of modulation frequencies can no longer be neglected. For such conditions, an extension of the ''single-channel'' model is required that allows integration of signal information across frequency. Such an extension of the single-channel model to a multi-channel model, that is able to simulate effects of spectral integration in amplitude-modulation detection and masking, is described in the accompanying paper (Dau *et al.*, 1997).

B. Modulation filterbank versus modulation low-pass filter

The modulation filterbank concept proposed here is different both from the previous version of the current model (cf. Dau *et al.*, 1996a, b) and from models proposed in the literature (e.g., Viemeister, 1979). Both of these models employed some kind of a modulation low-pass filter. The previous version of the current model was developed to describe simultaneous and forward masking data and included a low-pass filter with a cutoff frequency of 8 Hz. Such low-pass filtering, however, would fail to describe experiments concerning modulation masking and would also fail to describe basic experiments concerning modulation detection with narrow-band carriers at high center frequencies (cf. Sec. III A). The present model allows the prediction of modulation data and, at the same time, preserves the capabilities of the earlier model for describing simultaneous and nonsimultaneous masking data. This is because the linear modulation low-pass filtering (with a cutoff frequency at very slow modulations) is retained in the current model and is combined with the analysis of faster modulations by a modulation filterbank. The idea behind the modulation low-pass filter approach described in the literature (cf. Viemeister, 1979) is that a ''minimum integration time'' is typically derived from the cutoff frequency of the low-pass characteristic in the threshold function as a parameter that describes the auditory system's temporal resolution (for a review, see Viemeister and Plack, 1993). Such a model is capable of predicting a variety of different experiments, for example, the TMTF for broadband noise carriers (cf. Fig. 5) and modulation thresholds in third-octave band noise at different center frequencies (cf. Fig. 11). However, a model employing only a low-pass filter fails to describe the modulation detection data for a narrow-band noise carrier (cf. Fleischer, 1982a, 1983). Furthermore, such a model fails to describe the masking data using the tone-complex modulation masker. Hence, the model proposed here considerably expands the class of experiments that can be modeled correctly while still maintaining some of the properties and predictions of the model proposed by Viemeister (1979).

C. Intrinsic fluctuations of the noise carrier

In the past, only a few studies have attempted to involve the *inherent* statistical properties of the noise carriers in explaining and modeling TMTFs (e.g., Zwicker, 1953; Maiwald, 1967a, b; Fleischer, 1981, 1982b). For example, Fleischer (1981, 1982a, b) investigated TMTFs using narrow-band noise as the carrier. He developed a model for describing the interaction between inherent fluctuations within a noise carrier and the detectability of added modulation. The ''modulation spectrum'' was weighted by a certain factor which essentially represented a low-pass characteristic. For modulation frequencies lower than half the bandwidth of the noise carrier, this model yields good agreement with experimental results for modulation detection and modulation difference limens (Fleischer, 1981, 1982a, 1982b). For modulation frequencies larger than half the bandwidth of the noise carrier, this model would always predict a low-pass characteristic in the threshold function without regard to the applied carrier bandwidth—in the same way as shown in Sec. IV for the Viemeister model. Therefore, to account for the data, Fleischer extended the model by assuming ''cross-talk'' between the inherent fluctuations of the noise and the added modulation. He postulated a decay at a rate of 16 dB per decade of the modulation frequency to account for the high-pass characteristic in the data. In order to find a description for the inherent modulation of the noise, Fleischer (1981) regarded narrow-band noise with a bandwidth Δf as a pure tone which was amplitude modulated by a continuum of equal-amplitude modulation frequencies between zero and half the bandwidth of the noise. But this assumption is not correct. It would imply that the modulation spectrum of noise has a flat rather than a triangular shape, as shown by Lawson and Uhlenbeck (1950) (see the Appendix).

Even though the exact shape of the modulation spectrum assumed by Fleischer (1981) was not correct and the subsequent explanation of the data was based on a different concept than the one described here, Fleischer's concept of cross-talk between inherent envelope fluctuations of the carrier noise and the test modulation is compatible with the bandpass analysis proposed here. Within the modulation filterbank model, the low-pass characteristic of the threshold function for conditions with $f_{\mathrm{mod}} < \Delta f/2$ does not result from a specific weighing function used to model the ''sluggishness'' of the auditory system. Instead, it is a consequence of the intrinsic envelope fluctuations of noise bands and their spectral distribution on the one hand, and of the logarithmic scaling of the postulated modulation filters on the other hand. A critical test for this interpretation would be to obtain TMTFs for noise carriers with an envelope spectrum different from that of Gaussian noise, for example, low-noise noise (Hartmann and Pumplin, 1988; Kohlrausch *et al.*, 1997).

The current model can also account for the data using very narrow-band stimuli as the carrier, describing a high-pass or bandpass characteristic in the threshold function.

VI. CONCLUSIONS

(1) The experiments on modulation detection and modulation masking described here agree well with experiments from the literature. They provide a strong indication for an analysis of envelopes in terms of a separation into different modulation frequencies.

(2) The model of the effective signal processing in the auditory system proposed here is capable of quantitatively modeling most aspects of the experiments described. It employs a modulation filterbank for envelope analysis that exhibits a constant absolute bandwidth for low frequencies and a constant relative bandwidth for modulation frequencies above 10 Hz. Within the context of this model, the low-pass characteristic of the broadband TMTF is due to the inherent fluctuations of the carrier and constant relative bandwidth of the modulation filters, and not to a low-pass characteristic within the auditory system *per se*.

(3) While the predictions of the model proposed here agree with some predictions of the modulation low-pass model by Viemeister (1979) and an earlier version of the

current model (Dau *et al.*, 1996a, b), the model also holds for experiments such as modulation detection for narrow-band noise carriers, where the modulation low-pass approach clearly fails.

ACKNOWLEDGMENTS

We would like to thank all our colleagues of the Graduiertenkolleg Psychoakustik at the University of Oldenburg for fruitful discussions on the content of this paper. We also thank Brian Moore, Jesko Verhey, Andrew Oxenham, Stefan Münkner, and Ralf Fassel for their comments and suggestions concerning this study and for their critical reading of earlier versions of this paper. Two anonymous reviewers also provided very constructive criticism.

APPENDIX: ENVELOPE SPECTRA OF GAUSSIAN NOISES

The time-averaged power of the envelope is twice the average power of the waveform. Hence, it is independent of the noise bandwidth, as long as the total waveform power is fixed (cf. Hartmann and Pumplin, 1988). Therefore, for example, two Gaussian waveforms with the same power but with different bandwidths, have the same envelope power.

An interesting question is related to the *spectral* distribution of the envelope power. Lawson and Uhlenbeck (1950) calculated the spectrum of the envelope via Fourier transform of the autocorrelation function of the envelope. Assuming a rectangular shape of the power spectrum of the noise, they showed that the modulation spectrum $N=N(f_{\mathrm{mod}})$, i.e., the power spectrum of the (linear) envelope of the noise, is given approximately by the formula:

$$N_{\Delta f,\rho}(f_{\mathrm{mod}}) \approx \pi \Delta f \rho \, \delta(f_{\mathrm{mod}}) + \frac{\pi \rho}{4 \Delta f} \, (\Delta f - f_{\mathrm{mod}}), \qquad \text{(A1)}$$

where Δf is the noise bandwidth, ρ is the power spectral density, and f_{mod} indicates modulation frequency. Besides the dc peak represented by the δ function, an approximately triangular continuous spectrum results. In the case of the squared envelope, the modulation spectrum has *exactly* a triangular shape besides the dc peak. This corresponds to the Wiener–Chintchin theorem which states that the Fourier transform of the squared signal equals the autocorrelation of the spectrum of the signal.

The following aspects are of particular relevance for modulation-detection experiments using a narrow-band noise as carrier: For a constant overall level of a noise band, the total power of intrinsic noise fluctuations, i.e., the total area under the triangle, remains constant. What changes is the spectral region over which the envelope spectrum stretches. Hence, with increasing noise bandwidth, the modulation spectrum becomes broader and flatter.

[1]This modification was motivated by physiological studies on adaptation in auditory nerve fibres where a comparable ratio of onset and steady state response was found (e.g., Smith and Zwislocki, 1975; Westerman and Smith, 1984). It was further assumed that the too strong overshoot at the output of the adaptation model in its original version (see Dau *et al.*, 1996a, b) would have a detrimental effect on psychoacoustical threshold predictions. However, the limitation of the onset response by Münkner (1993) was found to not have a significant influence on the results in the present study.

[2]The transfer functions of the resonance filters can be derived from the following recursive function: $y_i = e^{-\pi B \Delta} \cdot e^{-i 2 \pi f_0 \Delta} \cdot y_{i-1} + (1 - e^{-\pi B \Delta}) \cdot x_i$, where B is the filter bandwidth, f_0 is the center frequency of the resonance filter, and Δ is the inverse sample rate. The output y_i at time i depends on the input x_i at time i and on the last output value y_{i-1}.

[3]The adaptation loops transform fast envelope fluctuations nearly linearly. However, in the framework of the present model, without any further nonlinearity at a level where the signal envelope has already been extracted, it would not be possible to simulate a sufficient amount of *masking* in conditions with random modulation maskers. Particularly, in such masking experiments, the scaling of the modulation filters would not have an effect and masked thresholds would not depend on signal modulation frequency which is in contrast with experimental data (see also the accompanying paper by Dau *et al.*, 1997). A physiological motivation for the calculation of the envelope of the modulation filter output may be given by the finding of Langner and Schreiner (1988) that a much greater percentage of neurons in the central nucleus of the inferior colliculus of the cat show sensitivity for modulation *rate* than for modulation phase, indicating that at this stage of processing a modulation-rate place coding is performed and modulation phase information is reduced. Such a coding has already been incorporated by Hewitt and Meddis (1994) in a computer model of amplitude-modulation sensitivity of single units in the IC.

[4]Because of the relatively broad tuning of the modulation filters, some energy of a (stationary) signal also leaks into the transfer range of the overlapping modulation filters tuned to ''higher'' modulation frequencies. Thus the internal representation contains signal information in parallel at the output of several modulation filters, whereas in the original model version (Dau *et al.*, 1996a, b) only the lowest modulation channel (low-pass) contributed to the decision. Therefore, in the corresponding calibration experiment, a somewhat higher variance of the internal noise at the output of each modulation filter is required to satisfy the 1-dB criterion compared to the variance adjusted with the modulation low-pass approach described in Dau *et al.* (1996a, b).

[5]The optimal detector realized in the model clearly is an *application* of the original concept of the optimal detector developed in signal detection theory by Green and Swets (1966), in which—for the case of signal known exactly—the signal itself (and not the signal with noise) is used for the correlation with the received signal. It should be noted, as already mentioned in Dau *et al.* (1996a), that in actual masking experiments, the signal is typically not presented in isolation, and that, second, the presence of the ''masker'' influences the internal representation of the signal in a nonlinear way. It appears to be an appropriate strategy to extract the internal representation of the (normalized) template at a level well above threshold—comparable with the situation at the beginning of an actual experiment—containing just a small amount of internal and external noise (see, for example, the template from Fig. 6 in the present study). It is further noted that the current modeling approach realizes a decision device acting at the level of the internal representation of the stimuli. All information about the signal that is available at this stage of processing, is used in an ''optimal'' way. That is, information is combined optimally, although, for example, modulation phase is lost at a certain stage of preprocessing.

Bacon, S. P., and Grantham, D. W. (**1989**). ''Modulation masking: Effects of modulation frequency, depth, and phase,'' J. Acoust. Soc. Am. **85**, 2575–2580.

Burns, E. M., and Viemeister, N. (**1981**). ''Played again SAM: Further observations on the pitch of amplitude-modulated noise,'' J. Acoust. Soc. Am. **70**, 1655–1660.

Creutzfeldt, O. D., Hellweg, F. C., and Schreiner, C. E. (**1980**). ''Thalamocortical transformation of responses to complex auditory stimuli,'' Exp. Brain Res. **39**, 87–104.

Dau, T. (**1996**). ''Modeling auditory processing of amplitude modulation,'' Doctoral thesis, University of Oldenburg.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). ''Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration,'' J. Acoust. Soc. Am. **102**, 2906–2919.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996a**). ''A quantitative model of the ''effective'' signal processing in the auditory system: I. Model structure,'' J. Acoust. Soc. Am. **99**, 3615–3622.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996b**). ''A quantitative model of the ''effective'' signal processing in the auditory system: II. Simulations and measurements,'' J. Acoust. Soc. Am. **99**, 3623–3631.

Eddins, D. (**1993**). ''Amplitude modulation detection of narrow-band noise: Effects of absolute bandwidth and frequency region,'' J. Acoust. Soc. Am. **93**, 470–479.
Fassel, R. (**1994**). ''Experimente und Simulationsrechnungen zur Wahrnehmung von Amplitudenmodulationen im menschlichen Gehör,'' Doctoral thesis, University of Göttingen.
Fassel, R., and Kohlrausch, A. (**1995**). ''Modulation detection as a function of carrier frequency and level,'' IPO Annual Progress Report **30**, 21–29.
Fassel, R., and Kohlrausch, A. (**1996**). ''Sinusoidal amplitude modulation thresholds as a function of carrier frequency and level,'' J. Acoust. Soc. Am. **99**, 2566.
Fassel, R., and Püschel, D. (**1993**). ''Modulation detection and masking using deterministic and random maskers,'' in *Contributions to Psychological Acoustics*, edited by A. Schick (Universitätsgesellschaft Oldenburg, Oldenburg), pp. 419–429.
Fassel, R., Kohlrausch, A., and Dau, T. (**1997**). ''The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers,'' submitted to J. Acoust. Soc. Am.
Fleischer, H. (**1981**). ''Amplitudenmodulation von Terzrauschen: Experimente und theoretische Ergebnisse,'' Acustica **47**, 155–163.
Fleischer, H. (**1982a**). ''Modulationsschwellen von Schmalbandrauschen,'' Acustica **51**, 154–161.
Fleischer, H. (**1982b**). ''Calculating psychoacoustic parameters of amplitude modulated narrow noise bands,'' Biol. Cybern. **44**, 177–184.
Fleischer, H. (**1983**). ''Modulation thresholds of narrow noise bands,'' Proceedings of the 11th ICA, Paris 1983, pp. 99–102.
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.
Goldstein, J. L. (**1967**). ''Auditory spectral filtering and monaural phase perception,'' J. Acoust. Soc. Am. **41**, 458–479.
Green, D. M. (**1973**). ''Temporal acuity as a function of frequency,'' J. Acoust. Soc. Am. **54**, 373–379.
Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).
Hartmann, W. M., and Pumplin, J. (**1988**). ''Noise power fluctuation and the masking of sine signals,'' J. Acoust. Soc. Am. **83**, 2277–2289.
Hewitt, M. J., and Meddis, R. (**1994**). ''A computer model of amplitude-modulation sensitivity of single units in the inferior colliculus,'' J. Acoust. Soc. Am. **95**, 2145–2159.
Holube, I., and Kollmeier, B. (**1996**). ''Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model,'' J. Acoust. Soc. Am. **100**, 1703–1716.
Houtgast, T. (**1989**). ''Frequency selectivity in amplitude-modulation detection,'' J. Acoust. Soc. Am. **85**, 1676–1680.
Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A., and Püschel, D. (**1997**). ''Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations,'' Acust. Acta Acust. **83**, 659–669.
Kollmeier, B., Dau, T., Hansen, M., and Holube, I. (**1996**). ''An Auditory-Model Framework for Psychoacoustics and Speech Perception and its Applications,'' Proceedings of the First Forum Acusticum, Antwerpen, published in Acust. Acta Acust. **82**, Suppl. 1, 89.
Langner, G. (**1992**). ''Periodicity coding in the auditory system,'' Hearing Res. **60**, 115–142.
Langner, G., and Schreiner, C. (**1988**). ''Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanism,'' J. Neurophysiol. **60**, 1799–1822.
Lawson, J. L., and Uhlenbeck, G. E. (**1950**). *Threshold Signals*, Volume 24 of *Radiation Laboratory Series* (McGraw-Hill, New York).
Levitt, H. (**1971**). ''Transformed up–down procedures in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Maiwald, D. (**1967a**). ''Ein Funktionsschema des Gehörs zur Beschreibung der Erkennbarkeit kleiner Frequenz- und Amplitudenänderungen,'' Acustica **18**, 81–92.
Maiwald, D. (**1967b**). ''Die Berechnung von Modulationsschwellen mit Hilfe eines Funktionsschemas,'' Acustica **18**, 193–207.
Martens, J. P. (**1982**). ''A new theory for multitone masking,'' J. Acoust. Soc. Am. **72**, 397–405.
Moore, B. C. J., and Glasberg, B. R. (**1986**). ''The role of frequency selectivity in the perception of loudness, pitch and time,'' in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London), pp. 251–308.
Münkner, S. (**1993**). ''Modellentwicklung und Messungen zur Wahrnehmung nichtstationärer Signale,'' Doctoral thesis, University of Göttingen.
Münkner, S., and Püschel, D. (**1993**). ''A psychoacoustical model for the perception of non-stationary sounds,'' in *Contributions to Psychological Acoustics*, edited by A. Schick (Universitätsgesellschaft Oldenburg, Oldenburg), pp. 121–134.
Palmer, A. R. (**1982**). ''Encoding of rapid amplitude modulations by cochlear-nerve fibers in the guinea pig,'' Arch. Otorhinolaryngol. **236**, 197–202.
Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1987**). ''An efficient auditory filterbank based on the gammatone function,'' in Paper presented at a meeting of the IOC Speech Group on Auditory Modelling at RSRE, December 14–15.
Püschel, D. (**1988**). ''Prinzipien der zeitlichen Analyse beim Hören,'' Doctoral thesis, University of Göttingen.
Ronken, D. A. (**1970**). ''Monaural detection of a phase difference in clicks,'' J. Acoust. Soc. Am. **47**, 1091–1099.
Schreiner, C., and Langner, G. (**1984**). ''Representation of periodicity information in the inferior colliculus of the cat,'' Soc. Neurosci. Abstr. **10**, 395.
Schreiner, C., and Urbas, J. V. (**1988**). ''Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields,'' Hearing Res. **32**, 49–65.
Sheft, S., and Yost, W. (**1990**). ''Temporal integration in amplitude modulation detection,'' J. Acoust. Soc. Am. **88**, 796–805.
Smith, R., and Zwislocki, J. (**1975**). ''Short-term adaptation and incremental responses of single auditory-nerve fibers,'' Biol. Cybern. **17**, 169–182.
Strickland, E. A., and Viemeister, N. F. (**1996**). ''Cues for discrimination of envelopes,'' J. Acoust. Soc. Am. **99**, 3638–3646.
Strickland, E. A., and Viemeister, N. F. (**1997**). ''The effects of frequency region and bandwidth on the temporal modulation transfer function'' J. Acoust. Soc. Am. **102**, 1799–1810.
Strube, H. W. (**1985**). ''A computationally efficient basilar-membrane model,'' Acustica **58**, 207–214.
Viemeister, N. F. (**1979**). ''Temporal modulation transfer functions based upon modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380.
Viemeister, N. F., and Plack, C. J. (**1993**). ''Time analysis,'' in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 116–154.
Wier, C., Jesteadt, W., and Green, D. M. (**1977**). ''Frequency discrimination as a function of frequency and sensation level,'' J. Acoust. Soc. Am. **61**, 178–184.
Westerman, L., and Smith, R. (**1984**). ''Rapid and short-term adaptation in auditory nerve responses,'' Hearing Res. **15**, 249–260.
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.
Zwicker, E. (**1953**). ''Die Veränderung der Modulationsschwellen durch verdeckende Töne und Geräusche,'' Acustica **3**, 274–278.
Zwicker, E. (**1956**). ''Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs,'' Acustica **6**, 365–381.

Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration[a)]

Torsten Dau[b)] and Birger Kollmeier
Carl von Ossietzky Universität Oldenburg, Graduiertenkolleg Psychoakustik, AG Medizinische Physik, D-26111 Oldenburg, Germany

Armin Kohlrausch
IPO Center for Research on User-System Interaction, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

(Received 28 June 1996; accepted for publication 4 August 1997)

A multi-channel model, describing the effects of spectral and temporal integration in amplitude-modulation detection for a stochastic noise carrier, is proposed and validated. The model is based on the modulation filterbank concept which was established in the accompanying paper [Dau *et al.*, J. Acoust. Soc. Am. **102**, 2892–2905 (1997)] for modulation perception in narrow-band conditions (single-channel model). To integrate information across frequency, the detection process of the model linearly combines the channel outputs. To integrate information across time, a kind of ''multiple-look'' strategy, is realized within the detection stage of the model. Both data from the literature and new data are used to validate the model. The model predictions agree with the results of Eddins [J. Acoust. Soc. Am. **93**, 470–479 (1993)] that the ''time constants'' associated with the temporal modulation transfer functions (TMTF) derived for narrow-band stimuli do not vary with carrier frequency region and that they decrease monotonically with increasing stimulus bandwidth. The model is able to predict masking patterns in the modulation-frequency domain, as observed experimentally by Houtgast [J. Acoust. Soc. Am. **85**, 1676–1680 (1989)]. The model also accounts for the finding by Sheft and Yost [J. Acoust. Soc. Am. **88**, 796–805 (1990)] that the long ''effective'' integration time constants derived from the data are two orders of magnitude larger than the time constants derived from the cutoff frequency of the TMTF. Finally, the temporal-summation properties of the model allow the prediction of data in a specific temporal paradigm used earlier by Viemeister and Wakefield [J. Acoust. Soc. Am. **90**, 858–865 (1991)]. The combination of the modulation filterbank concept and the optimal decision algorithm proposed here appears to present a powerful strategy for describing modulation-detection phenomena in narrow-band and broadband conditions.
[S0001-4966(97)05711-1]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Mk [JWH]

INTRODUCTION

One of the most interesting and fundamental questions in psychophysical research is how the auditory system ''trades'' spectral and temporal resolution. One problem in this field is the question of how peripheral filtering affects the ability to detect modulation. It is often postulated that peripheral filtering does not limit the ability to detect modulation and that in ''temporal resolution'' tasks such as modulation detection the observer broadens his ''effective'' bandwidth (e.g., Viemeister, 1979; Berg, 1996). With such an assumption, the experimental data can be simulated by a model using a ''predetection'' filter as broad as several critical bandwidths to account for some kind of peripheral filtering (Viemeister, 1979). If, in contrast, peripheral filtering in terms of critical bands has any influence on modulation detection, the question arises how the applied modulation analysis depends on center frequency. In this vein, some authors have postulated that the TMTF of a broadband noise is actually determined by the information in the *highest* critical band excited by the stimulus (e.g., Maiwald, 1967a, b; Van Zanten, 1980). Eddins (1993) examined spectral integration in amplitude-modulation detection, independently varying stimulus bandwidth and spectral region. He found that the cutoff frequency of the TMTF does not depend on carrier frequency region, but increases significantly with increasing carrier bandwidth. As recently pointed out by Strickland and Viemeister (1997), this latter observation might be an artifact of the stimulus generation, which included a rectangular bandlimitation after modulation of the carrier.

In Dau (1996) and Dau *et al.* (1997), experiments on amplitude-modulation detection were described using narrow-band noise as the carrier at a high center frequency (5 kHz). By using these conditions, effects of any peripheral or ''predetection'' filtering were minimized and, in addition, the spectral region that was being used to detect the modulation was restricted to one critical band. A model of the effective signal processing in the auditory system, including a modulation filterbank, was derived to analyze the temporal envelope of the stimuli. This model will be called the

[a)]Part of this research was presented at the 131th meeting of the Acoustical Society of America [T. Dau, B. Kollmeier and A. Kohlrausch, ''A quantitative prediction of modulation masking with an optimal-detector model,'' J. Acoust. Soc. Am. **99**, 2565(A) (1996)].

[b)]Corresponding author. Electronic mail: torsten@medi.physik.uni-oldenburg.de

''modulation filterbank model'' throughout this paper.

To get more insight into the processing of modulation, several experiments are described here that investigate the effects of spectral integration in amplitude-modulation detection, examining in particular the transition between narrow-band and broadband conditions. To test the capabilities of the modulation filterbank model in conditions of spectral integration, some critical experiments are performed and model predictions are compared with data from the literature (Eddins, 1993; Sheft and Yost, 1990) and the new experimental data. To compare results from experiments and simulations with as close a similarity in the experimental details as possible, the experiments by Eddins (1993) and those by Sheft and Yost (1990) were replicated in our laboratory with a slightly different threshold estimation procedure.

Another aspect of modulation perception is the phenomenon of temporal integration in amplitude-modulation detection. Temporal integration or temporal summation refers to the well-known fact that over a range of durations thresholds decrease with increasing signal duration. Several models have been suggested in the literature to describe this phenomenon. However, differences in the modeling strategies occur that reflect the ''resolution-integration conflict'' (de Boer, 1985). In the case of temporal integration in modulation detection, Sheft and Yost (1990) found that time constants associated with temporal integration are much larger than those indicated by the ''resolution data.'' The present study examines the ability of the modulation filterbank model to account for these long effective time constants found in the data.

As before, both our own experimental data from ''critical'' experiments as well as data from the literature were compared with model predictions. It should be emphasized at this point that the same parameter set was used for all model predictions throughout this paper. These parameters and in particular the scaling of the modulation filters were also the same as in the accompanying paper (Dau *et al.*, 1997), where the filterbank parameters were adjusted to predict amplitude modulation-detection thresholds for narrow-band noise carriers.[1]

I. MULTI-CHANNEL MODEL

All simulations discussed here were performed on the basis of the modulation filterbank model that was initially developed as a single-channel model, as described in Dau *et al.* (1997). Figure 1 shows the structure of the multi-channel model which was extended from the single-channel model by performing the modulation analysis in parallel on the output of each stimulated peripheral auditory filter. This is motivated in part by results from physiological studies of the representation of amplitude modulation within the inferior colliculus (IC) of the cat, where it was found that modulation frequencies are represented in a systematic way orthogonally to the tonotopical organization of the IC (Schreiner and Langner, 1988).

The model contains the same stages of signal processing along the auditory pathway as proposed in the single-channel model. The main features are described briefly here; for further details see Dau *et al.* (1996a, 1997). To simulate the

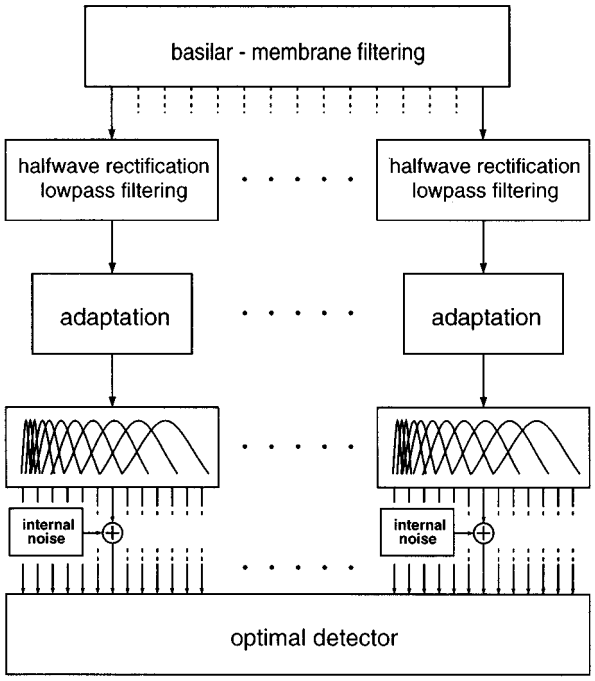


FIG. 1. Block diagram of the psycho-acoustical model for describing modulation detection data with an optimal detector as the decision device. The signals are preprocessed, subjected to adaptation, filtered by a modulation filterbank and finally added to internal noise; this processing transforms the signals into their ''internal representations.''

peripheral bandpass characteristics of the basilar membrane, the gammatone filterbank of Patterson *et al.* (1987) is used. The filters overlap at the -3-dB points of their transfer functions. The summation of all filter transfer functions would result in a flat transfer characteristic across frequency. The stimulus at the output of each peripheral filter is half-wave rectified and low-pass filtered at 1 kHz. This stage essentially preserves the envelope of the signal for high center frequencies. Effects of adaptation are simulated by nonlinear adaptation circuits (Püschel, 1988; Dau *et al.*, 1996a). The parameters for these circuits were the same as in Dau *et al.* (1996a, b). The adapted signal is further analyzed by a linear modulation filterbank. The parameters of the modulation filters agree with those in Dau *et al.* (1997) and we refer to that paper for a detailed description. It is assumed within the present study that these parameters do not vary across frequency, i.e., the same modulation filterbank is applied to analyze the signal's envelope at the output of each critical band. Limitations of resolution are simulated by adding internal noise with a constant variance to each modulation filter output. The decision device is realized as an optimal detector in the same way as described in Dau *et al.* (1996a, 1997). Within the multi-channel model, the internal representation of the stimuli has four dimensions, namely amplitude, time, modulation center frequency (as before in the single-channel model), and (audio) center frequency as the additional axis. In the simulations the internal representations of the different peripheral channels are appended one after another in one large array. The decision process works exactly in the same way as described in the single-channel analysis, with the extension that all auditory filters with cen-

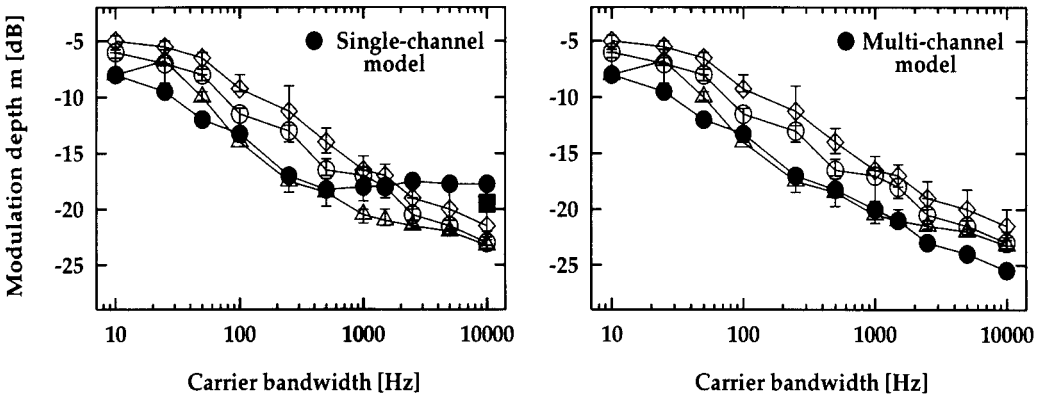


FIG. 2. Modulation-detection thresholds of a 5-Hz modulation as a function of the carrier bandwidth. Open symbols indicate measured data of three subjects (same in the left and the right panel). Simulated thresholds with the single-channel model are represented in the left panel by the filled circles. Simulated thresholds with the multi-channel model are shown in the right panel (filled circles). The filled box represents the simulated threshold for $\Delta f = 10$ kHz, where the modulation analysis was performed using the highest critical band excited by the stimulus. Center frequency: 5 kHz; Carrier and modulation duration: 500 ms. Level: 65 dB SPL. Subjects: JV (□); TD (○); PD (◇); optimal detector (●,■).

ter frequencies within the spectral range of the stimulus were included in the analysis. Decisions were then based on the cross correlation between the (four-dimensional) actual internal representation of the stimulus and the normalized suprathreshold (four-dimensional) template representation. Using such a model, a prediction of the average subject's performance is possible on a trial-by-trial basis. Simulations of the experimental runs can thus be performed by implementing on the computer the same threshold estimating procedures within the model as those being used in the experiments with human observers.

II. METHOD

Modulation thresholds were measured and simulated using an adaptive 3-interval forced-choice (3IFC) procedure. The experimental procedure and stimulus generation were the same as described in the accompanying paper (Dau *et al.*, 1997). Also, the same five subjects participated in this study.

Several experiments on modulation detection were performed. The subject's task was to detect a sinusoidal test modulation of a Gaussian noise carrier. The test modulation was applied with zero onset phase. Unless explicitly stated, a carrier level of 65 dB SPL was used. During the experiments with a noise carrier, an independent sample of the noise was presented in each interval. Filtering of the noise stimuli was done in the frequency domain by Fourier transforming the whole noise waveform, setting the respective frequency samples to zero, and transforming the signal back into the time domain. Unless stated otherwise, the filtering was applied prior to modulation. The rms level of the modulated signal was always equated to the rms level employed in the unmodulated trials.

III. RESULTS

A. Modulation analysis within and beyond one critical band

In the first experiment, the effect of carrier bandwidth on the detectability of a certain signal modulation was examined for bandwidths within and beyond one critical band. In the accompanying study by Dau *et al.* (1977) only carrier bandwidths smaller than a critical band were applied and in the corresponding simulations the modulation analysis was performed only in one peripheral channel. The present experiment was designed to illustrate the limitations of the ''single-channel'' model in conditions of spectral integration in modulation detection, and to show the necessity of the extention of the model towards a ''multi-channel'' model.

A noise carrier centered at 5 kHz was sinusoidally modulated with a modulation rate of 5 Hz. Modulation depth at threshold was measured as a function of the carrier bandwidth, which had one of the following values: 10, 25, 50, 100, 250, 500, 1000, 1500, 2500, 5000, or 10 000 Hz. The duration of the carrier and the modulation was 500 ms including 50-ms cosine-squared rise and decay ramps. The bandwidths cover the range from less than the critical bandwidth at 5 kHz to much greater than the critical bandwidth. The left panel of Fig. 2 shows the experimental results of three subjects (open symbols). The ordinate indicates modulation depth at threshold and the abscissa shows the carrier bandwidth. Thresholds decrease monotonically with increasing bandwidth. This decrease is due to changes in the spectrum of the inherent fluctuations, which, for low modulation frequencies, decreases with increasing bandwidth of the carrier (Lawson and Uhlenbeck, 1950; see Dau *et al.*, 1997). As a consequence, less energy of the random envelope fluctuations of the carrier leaks into the transfer range of the modulation filters near the test modulation frequency. In the model, this leads to decreasing thresholds with increasing carrier bandwidth. Simulated thresholds obtained with the single-channel model as described in Dau *et al.* (1997) are plotted in the left panel of Fig. 2 (filled symbols). In this single-channel simulation, the modulation analysis was carried out using the peripheral filter tuned to 5 kHz. In conditions with carrier bandwidths $\Delta f < 1000$ Hz, simulated and measured data are in good agreement. For bandwidths $\Delta f \geqslant 1000$ Hz, however, systematic differences between simulations and experimental data occur, which increase with in-

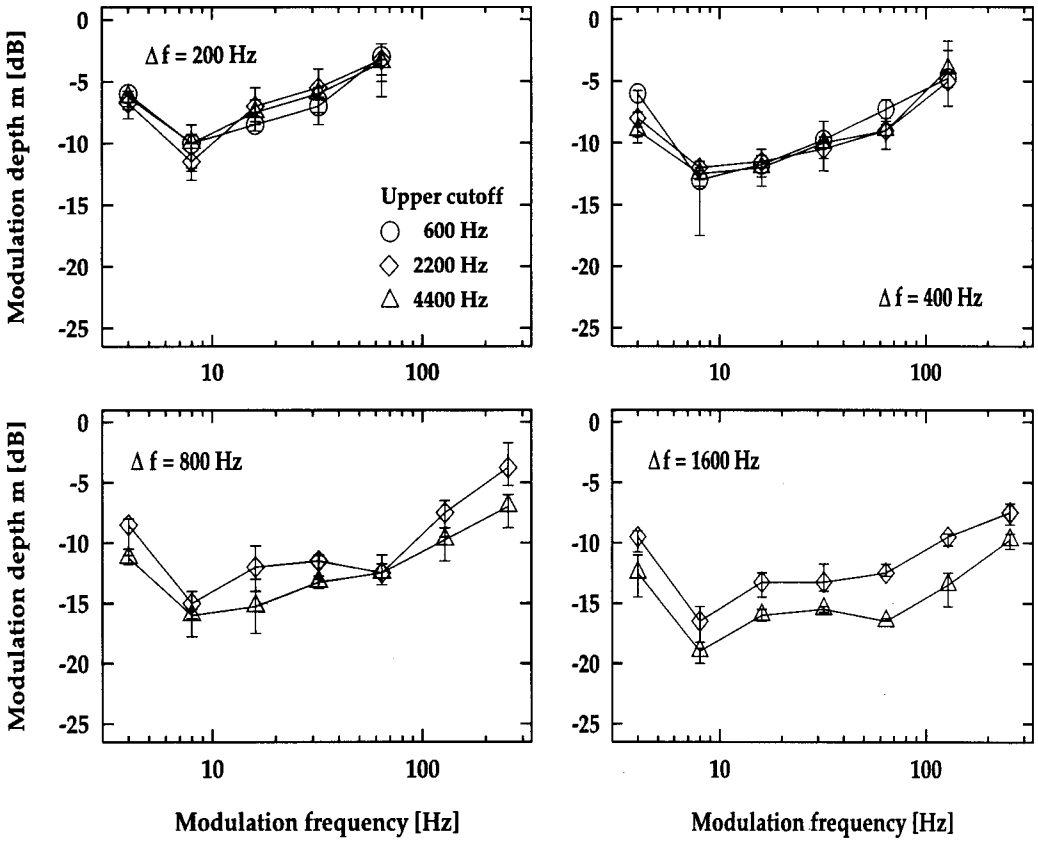


FIG. 3. Modulation-detection thresholds for one subject (BG) as a function of modulation frequency with upper cutoff frequency as the parameter. Each panel represents a different bandwidth condition: upper left: 200 Hz, upper right: 400 Hz, lower left: 800 Hz, lower right: 1600 Hz. The upper cutoff frequency is either 4400 Hz (△); 2200 Hz (◇); or 600 Hz (○).

creasing bandwidth. For the bandwidth $\Delta f = 10$ kHz, for example, the difference between the experimental results and simulations amounts to more than 5 dB. Of course, this discrepancy is expected, since the simulations were carried out using only the information about the signal modulation available at the output of one peripheral filter. The filled box in the left panel of Fig. 2 represents the simulated threshold for a carrier bandwidth of 10 kHz, where the modulation analysis was performed using the highest critical band excited by the stimulus. Even in this ''optimized'' single-channel simulation, there remains a discrepancy of nearly 5 dB between the simulated and the measured threshold. This does not support the hypothesis by Maiwald (1967a, b) and Van Zanten (1980) that sufficient information about the signal modulation is available in the highest excited frequency region and that the detection strategy of the subject is to monitor the ''internal'' filters in this high-frequency region.[2] The right panel of Fig. 2 shows the simulated data from the multichannel filterbank model together with the experimental data already shown in the left panel. By combining information from all excited peripheral channels, the model can account for the continuing decrease in thresholds over the whole range of carrier bandwidths.

B. Effects of bandwidth and frequency region

In the next experiment, effects of absolute bandwidth and frequency region on spectral integration in modulation detection were studied in the same way as by Eddins (1993). Modulation thresholds for narrow-band noise carriers were measured as a function of modulation frequency for the following conditions: The stimulus bandwidth was either 200, 400, 800, or 1600 Hz and the frequency region was varied by adjusting the high-frequency cutoff of the noise to be either 600, 2200, or 4400 Hz. The purpose was to determine TMTFs by *independently* varying stimulus bandwidth and stimulus frequency to determine the influence of these parameters on modulation detection. As in Eddins (1993), the modulated stimuli were generated by bandpass filtering *after* amplitude modulation of wide-band noise to avoid the possibility that the detection of modulation would be based on spectral cues in the signal interval rather than changes in the temporal waveform. In this way, the bandwidth of the narrow-band stimuli was the same in the presence or absence of modulation. In addition, the stimuli were adjusted to have equal energy in each interval of the forced-choice trial to prevent detection of modulation based on overall level rather than on the presence or absence of modulation.

Figures 3 and 4 show the experimental data of two subjects for several stimulus conditions. Modulation depth *m* at threshold is plotted as a function of modulation frequency. The transfer functions reflect a bandpass characteristic that is similar to data from previous studies (Rodenburg, 1972, 1977; Viemeister, 1977, 1979; Formby and Muir, 1988). The data are in very good agreement with the results of Eddins,

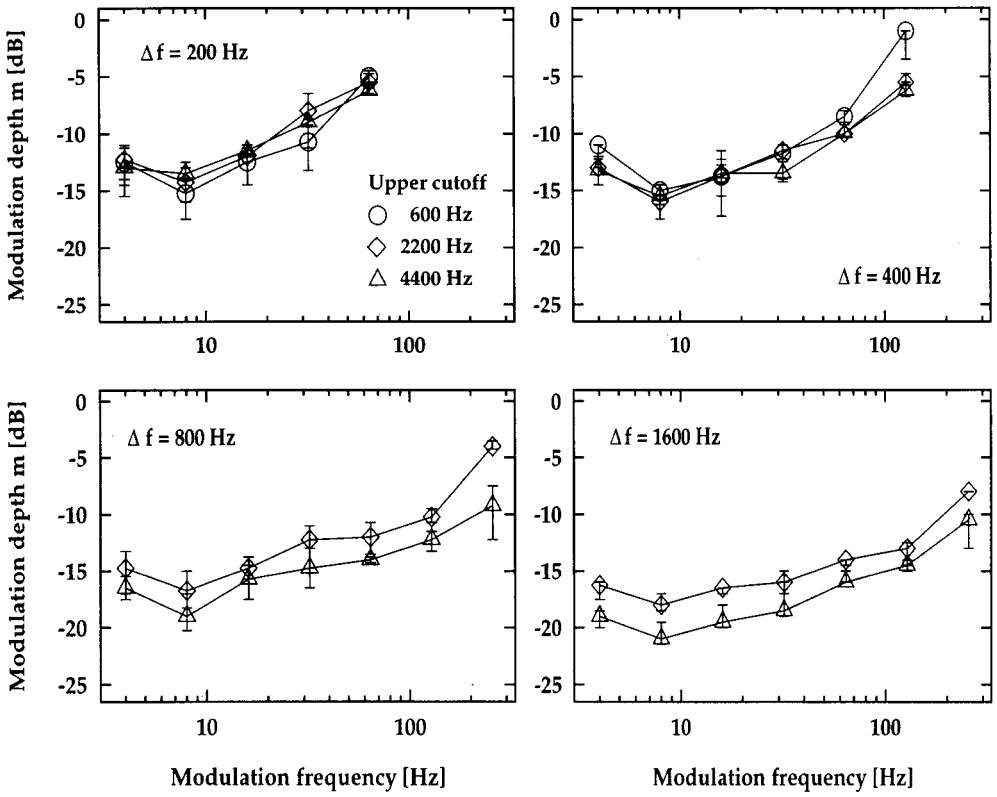


FIG. 4. Same as Fig. 3 for subject TD.

in spite of slight differences in the threshold estimation procedure between both studies. Increasing stimulus bandwidth results in a corresponding increase in sensitivity to modulation. This systematic increase is seen for each of the three cutoff frequencies tested. Eddins fitted his data—for each subject and condition—with the transfer function of a simple low-pass filter and derived the time constant associated with the -3-dB cutoff frequency of the specific transfer function. Two interpretations emerged from his analysis: First, the time constants associated with the TMTFs do not vary with changing frequency region. Eddins (1993) concluded from the data that temporal acuity is independent of frequency region, assuming that temporal acuity and derived time constants from the data are directly related. Second, the time constants associated with TMTFs decrease monotonically with increasing stimulus bandwidth.

Figure 5 shows the corresponding model predictions obtained with the modulation filterbank model. Modulation depth at threshold is plotted as a function of modulation frequency. Modulation analysis was performed using those peripheral channels with center frequencies within the spectral width of the noise carrier. The shape of the simulated threshold patterns agrees well with the corresponding data, but there is a constant absolute deviation of 2–5 dB between model predictions and individual experimental data. That is, the *form* of the simulated TMTF does not depend on the frequency region of the stimuli but depends on the stimulus bandwidth. In some—extreme—conditions, e.g., for $\Delta f = 400$ Hz at the upper cutoff frequency of 600 Hz and for $\Delta f = 800$ Hz at the upper cutoff frequency of 2200 Hz, the slope of the simulated TMTF between the penultimate and the last modulation frequency is steeper than in case of other cutoffs at these bandwidths. The same tendency is seen in most of the corresponding experimental conditions, e.g., in Fig. 3 for 800 Hz and in Fig. 4 for 400 and 800 Hz.

Within the model the ''low-frequency'' and the ''high-frequency'' conditions require quite different decision strategies to determine the detection threshold. Consider, for example, the bandwidth of $\Delta f = 400$ Hz and the upper cutoff frequencies of 600 and 4400 Hz, representing the low-frequency and the high-frequency conditions, respectively. In the low-frequency condition, the modulation analysis is performed in parallel in several peripheral filters. In the high-frequency condition, the modulation analysis is carried out in only one peripheral channel because of the poor spectral resolution of the auditory system at high frequencies. The results of the simulations in the modulation detection task are the same for both conditions (except for the highest modulation frequency), indicating that, apparently, the poor temporal resolution in the low-frequency condition is somehow compensated for by the greater number of ''observations'' across frequency compared to the high-frequency condition. This compensation will break down at high modulation frequencies, because these will be strongly attenuated, if many narrow auditory filters cover the carrier spectrum. In this case, a second limitation for detecting modulation comes into play, namely the *absolute threshold* for modulation detection. Such a threshold was introduced within the present model by the addition of *internal* noise to the output of each modulation filter. For a sinusoidal carrier this internal noise

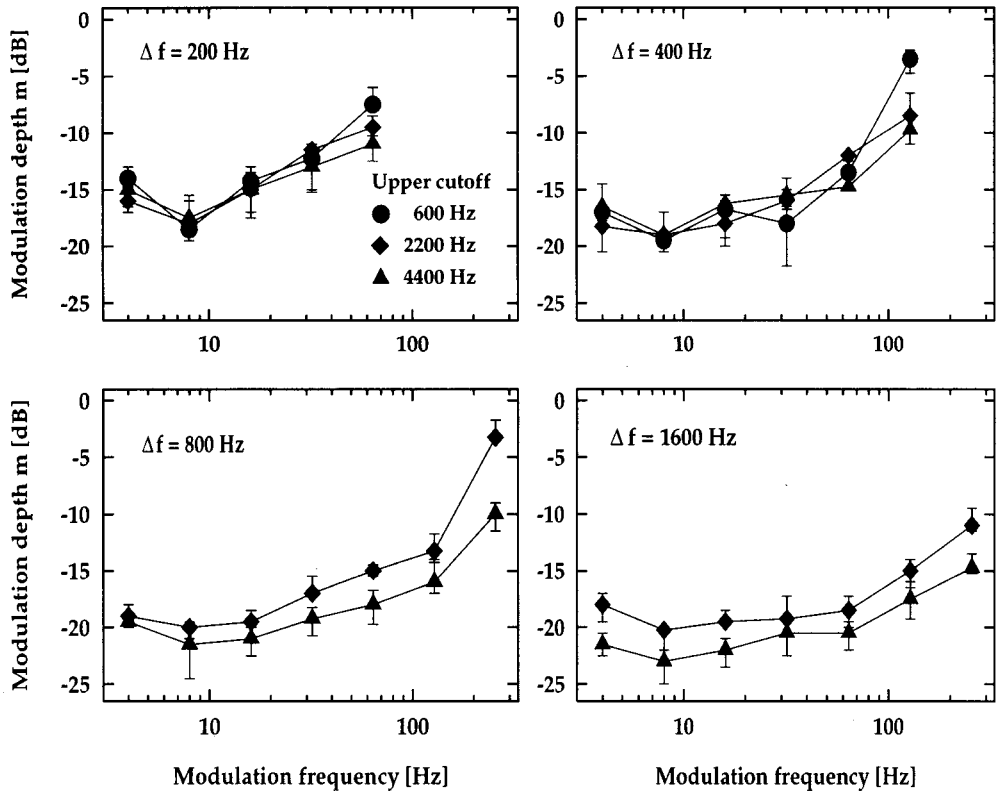


FIG. 5. Simulated modulation-detection thresholds as a function of modulation frequency for the same conditions as in Figs. 3 and 4. Thresholds are indicated by filled symbols.

results in a threshold of detectability for a low-frequency amplitude modulation of about −27 dB. If, in the case of a noise carrier, the imposed modulation is strongly attenuated within one or more peripheral filters, the detectability of the modulation is not determined solely by the inherent statistics of the noise carrier but is also determined by the internal noise. As a consequence, in such a condition, some of the (low-frequency) peripheral filters do not contribute to the information about the signal modulation leading to an increased detection threshold. Therefore, in an ''extreme'' condition as described above, namely in the case of the highest imposed modulation frequency in the lowest-frequency region, the width of the peripheral filter has an influence on modulation detection. This influence is indeed seen in the model predictions and in the experimental data for some, but not all, subjects.

C. Influence of filter shape and spacing on spectral integration

In some additional simulations we investigated the influence of filter shape and filter spacing on spectral integration. These simulations were performed for a low modulation frequency of 8 Hz and involved high-frequency auditory channels of a broadband running noise carrier. Thresholds are based on 20 repeated estimates of the model.

The gain resulting from the analysis of coherent modulation in several filters is a consequence of combining (partially) independent observations. As long as external fluctuations are the limiting factor for detection, the independence of these observations will depend on the correlation between the carrier waveforms in these channels. The correlation depends in turn on the shape and spacing of the filters. In a first simulation, we implemented nonoverlapping rectangular filters in the model instead of filters with a Roex shape. In this situation, the carrier waveforms in the various filters are completely independent. This condition results in a threshold decrease with 3 dB per doubling of the number of analyzed filters. This is the same amount which will later be described and discussed in the context of temporal integration (see Sec. IV B). In a similar simulation with Roex filters overlapping as in the present model, the gain was somewhat smaller and amounted to about 2.5 dB per doubling of the number of filters (8-dB threshold change by going from one to nine filters). This result shows that the remaining correlation between adjacent Roex filters has a small, but ''measurable'' effect on the threshold prediction.

In a second simulation, we compared predictions for a single high-frequency filter with those for the combination of five adjacent Roex filters. These five filters either had a spacing as in the present model (narrow spacing) or they were separated twice as much (wide spacing). In agreement with the previous simulation, the combination of narrower spaced filters resulted in a threshold 5.5 dB below the prediction for a single filter. In the wide-spaced condition, thresholds were an additional 1.25 dB lower. Compared with the single-filter result, the combination of widely spaced Roex filters led to a 3-dB effect per doubling of the number of filters. We thus can conclude that by allowing some overlap between adja-

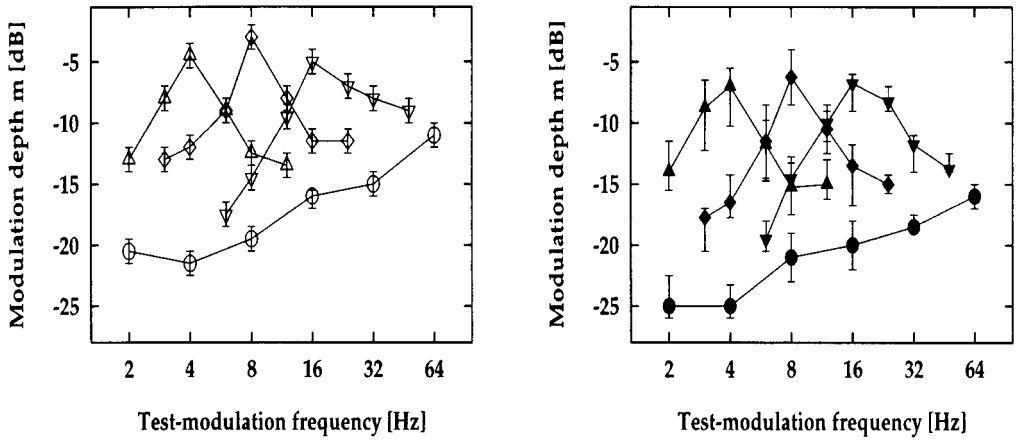


FIG. 6. Modulation thresholds for a sinusoidal test modulation and a pink noise ranging from 1 to 4 kHz as carrier. The lower curve in both panels shows the unmasked modulation detection thresholds as a function of modulation frequency. Each of the three peaked curves show the masked-modulation threshold pattern for one of the $\frac{1}{2}$-octave-wide masker-modulation bands centered at 4 Hz (△), 8 Hz (◇), or 16 Hz (▽). The left panel shows data from Houtgast (1989). The right panel shows predictions with the multi-channel model.

cent filters, the observations in these filters are not statistically independent. By using a wider spacing, the information gain from combining a certain number of filters is larger. However, with a wider spacing, the number of filters that could be placed within a *given* spectral range would decrease. This decrease in the number of observations would affect thresholds more strongly than the gain from the statistical independence, so that the spacing used in the present model makes better use of the available information than a wider spacing would.

Finally, we checked whether an even narrower spacing of filters would change the model predictions. This was done for one of the conditions from Eddins (1993) described in the previous section. The carrier was a bandlimited noise ranging from 2800 to 4400 Hz. In the first simulation, modulation analysis was performed with Roex filters overlapping at their −3-dB points. In a second simulation, the spectral density and thus the total number of analyzed filters was increased by a factor of 3. Both simulations led to the same threshold prediction.

In summary, the additional simulations show that the chosen overlap between adjacent filters allows a close to optimal detection of modulation imposed on a broadband noise carrier with a minimal computational load (i.e., the smallest number of analyzed auditory channels).

D. Predictions for modulation masking using broadband noise carriers

Houtgast (1989) adopted a classical masking paradigm for investigating frequency selectivity in the modulation-frequency domain: the detectability of test modulation in the presence of masker modulation, as a function of the spectral difference between test and masker modulation.

The carrier in all his experiments was a pink noise with a spectrum level in the 1-kHz region of about 25 dB SPL. After applying the modulation, the carrier was bandpass filtered between 1 and 4 kHz and added to unmodulated pink noise with a complementary (bandstop) spectrum. The masking patterns (of the first experiment) were obtained for each of three $\frac{1}{2}$-octave-wide bands of noise as the modulation masker. The carrier noise was multiplied both with the masker and the target modulator waveform. Center frequencies for the masker modulation were 4, 8, or 16 Hz. The left panel of Fig. 6 shows the resulting masking patterns from that study. The three curves show a peaked characteristic: The amount of masking decreases for increasing frequency difference between test modulation and masker modulation. The lower curve in the figure shows the unmasked modulation detection threshold level as a function of modulation frequency. For details about the experimental setup, stimuli and procedure, see Houtgast (1989).

The right panel of Fig. 6 shows results from simulations obtained with the multi-channel version of the modulation filterbank model. Instead of using notched noise surrounding the carrier, we forced the model to only analyze auditory filters with center frequencies between 1 and 4 kHz, which should have the same effect with respect to off-frequency listening. The simulated thresholds for the unmasked condition are systematically lower than the experimental data from Houtgast. This might be caused by the different presentation of the masker and by differences in the applied threshold estimation procedure. Besides this systematic difference of about 2–5 dB, both the unmasked thresholds and the masked threshold in both panels agree very well with each other. The simulated masked patterns show the same peaked characteristic, with less masking for increasing frequency difference between test and masker modulation.

An additional observation can be made from the unmasked modulation threshold data in Fig. 6. The difference between the experimental thresholds at 4 and 64 Hz amounts to about 11 dB. This is much more than observed in corresponding measurements with a white-noise carrier (cf. Figs. 3 and 4 in Sec. III B). Interestingly, a similar difference between white and pink noise is also observed in simulations. When the experimental condition of Houtgast is simulated with a white-noise carrier, thresholds at 4-Hz modulation are the same as for pink noise. The increase between 4 and 64 Hz is, however, smaller for white noise and amounts to about

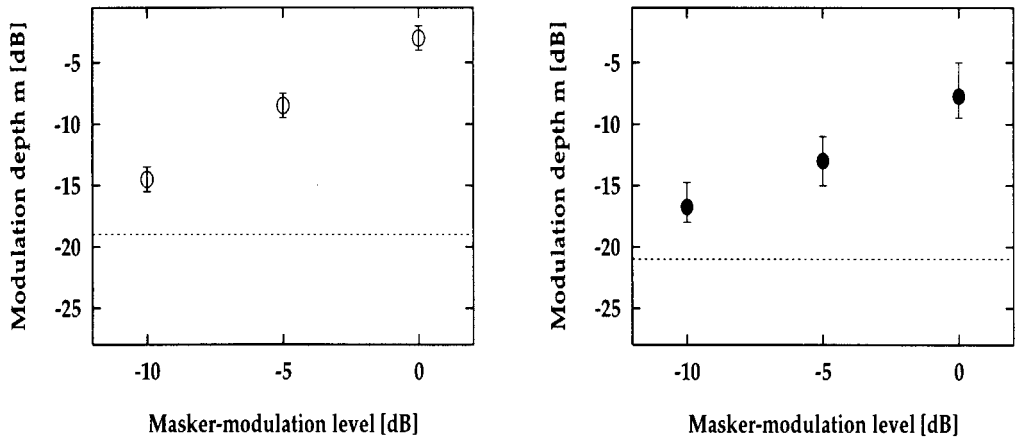


FIG. 7. 8-Hz modulation-detection thresholds as a function of the relative masker-modulation level with the masker-modulation band centered at 8 Hz. Left panel: data from Houtgast (1989). Right panel: corresponding predictions of the modulation filterbank model. The dashed lines represent the unmasked detection threshold of the test modulation.

7 dB. At present there is no simple explanation for this influence of the masker's spectral shape on the TMTF.

As in the study of Houtgast (1989), the effect of a reduction of the masker-modulation level was investigated for one of the conditions from Fig. 6 (the 8-Hz modulation frequency for the noise band centered at 8 Hz). The masker-modulation level used to obtain the middle threshold function in Fig. 6 was reduced by 5 and 10 dB, respectively. The effect on the model prediction is shown in Fig. 7 (right panel). The reduction of the threshold level amounts to 9 dB and is slightly smaller than the reduction of the masker-modulation level (10 dB) and 2 dB smaller than the effect observed in the data by Houtgast (11 dB).

To further test the ability of the modulation filterbank model to account for modulation masking data, the effect of varying the bandwidth of the modulation masker was investigated. Thresholds were obtained for a test modulation frequency of 8 Hz for various values of the bandwidth of the masker modulation. The center frequency and the spectral density within the passband were kept constant. Figure 8 shows results from the study of Houtgast (1989) (left panel) and model predictions obtained with the present model (right panel).

Houtgast (1989) found that for small bandwidths, the modulation thresholds increased by 3 dB for each doubling of the masker-modulation bandwidth, whereas for large bandwidths, the threshold remained constant. He proposed that the modulation-detection threshold is associated with a constant signal-to-noise ratio within a filter centered on the test-modulation frequency. To a first approximation, indicated by the two straight lines in the left panel of Fig. 8, he suggested a width of the modulation filter of $\frac{1}{2}$–1 oct.[3]

In the present model, the modulation filter centered at 8 Hz has a bandwidth of 5 Hz. This bandwidth lies exactly in the range of $\frac{1}{2}$–1 octave that was suggested by Houtgast. The model predicts an increase in threshold of about 6 dB between $\frac{1}{8}$ and $\frac{1}{2}$ octave of modulation-masker bandwidth. For modulation-masker bandwidths larger than a $\frac{1}{2}$ oct, it predicts an almost constant threshold. This is in agreement with the data of Houtgast. Thus modulation masking is only effective within a ''critical'' band around the test-modulation frequency.

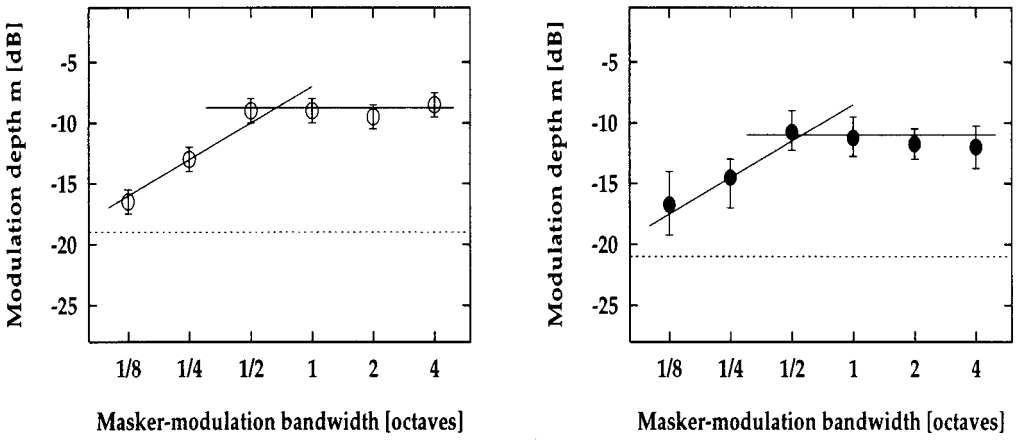


FIG. 8. 8-Hz modulation-detection thresholds as a function of the bandwidth of the masker-modulation noise centered at 8 Hz. Left panel: data from Houtgast (1989). Right panel: corresponding predictions of the modulation filterbank model.

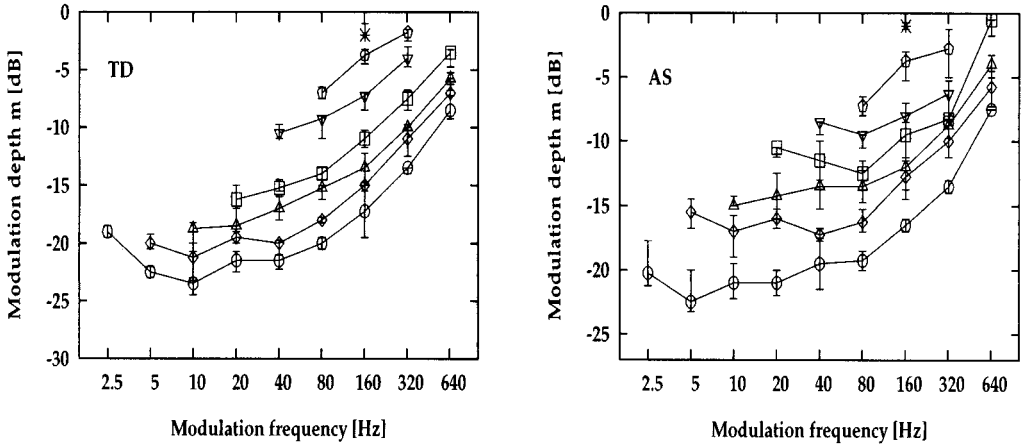


FIG. 9. Measured TMTFs of two subjects using broadband noise as a carrier. Parameter is the duration of the imposed test modulation. Carrier: 0–6 kHz; Level: 65 dB SPL; Modulation duration: ○: 400 ms; ◇: 200 ms; △: 100 ms; □: 50 ms; ▽: 25 ms; ⬡: 12.5 ms; ×: 6.25 ms; Subjects: TD (left panel); AS (right panel).

The results from these simulations support the hypothesis that the envelope fluctuations of the stimuli are processed by modulation-frequency selective channels.

E. Temporal integration in modulation detection

Thresholds for detecting sinusoidal amplitude modulation of a wide-band noise carrier (low-pass filtered at 6 kHz) were measured and simulated as a function of the duration of the modulating signal. The experimental design was chosen according to the study of Sheft and Yost (1990). The carrier was gated with a duration that exceeded the duration of modulation by the combined stimulus rise and fall times. Stimuli were shaped by a 25.6-ms rise–fall time. The combinations of modulation frequency and numbers of modulation cycles for each condition are listed in Table I in Sheft and Yost (1990). Because modulation was restricted to the constant-amplitude portion of the carrier (thus excluding the ramps), stimulus duration was always 51.2 ms longer than the modulation duration listed in the table (see Sheft and Yost, 1990).

Figure 9 shows the experimental data of two subjects. The modulation depth at threshold is plotted as a function of the modulation frequency. The parameter is the modulation duration. The data agree very well with those from Sheft and Yost (1990). The curves represent the ''classical'' broadband TMTF often described in other studies (e.g., Viemeister, 1979; Formby and Muir, 1988). The data show somewhat increased thresholds for the two lowest modulation frequencies, $f_{\mathrm{mod}}=2.5$ and 5 Hz. This is caused by the ''gating'' of the carrier in this experiment in contrast to experiments where the carrier was presented continuously, as discussed in previous studies (e.g., Viemeister, 1979; Sheft and Yost, 1990).[4] For modulation frequencies between 5 and 40 Hz, the thresholds are roughly constant. They increase slightly between 40 and 80 Hz and at a rate of approximately 3 dB/octave for higher modulation frequencies.

Figure 10 shows the corresponding simulated thresholds obtained with the modulation filterbank model. The ordinate indicates modulation depth at threshold and the abscissa represents modulation frequency. The simulated threshold pattern is in very good agreement with the pattern found in the measured data. It shows increased thresholds for the two lowest modulation frequencies, 2.5 and 5 Hz, leading to a bandpass characteristic of the threshold function for the two greatest durations of 200 and 400 ms. This is caused by the dynamic properties of the adaption model (Püschel, 1988; Dau *et al.*, 1996a). The feedback loops of the adaptation model produce a considerable overshoot at the carrier onset that decreases the detectability of the signal modulation especially at very low modulation rates (see discussion in Viemeister, 1979; Sheft and Yost, 1990).[5]

The time constants derived from temporal integration per unit duration are in very good agreement with those found in the measured data. The model therefore accounts for the long effective integration time constants whose values are much larger than the time constants indicated by the ''resolution data'' (Sheft and Yost, 1990; Viemeister, 1979).

IV. DISCUSSION

In this study the performance of the modulation filterbank model was described with respect to spectral and tem-

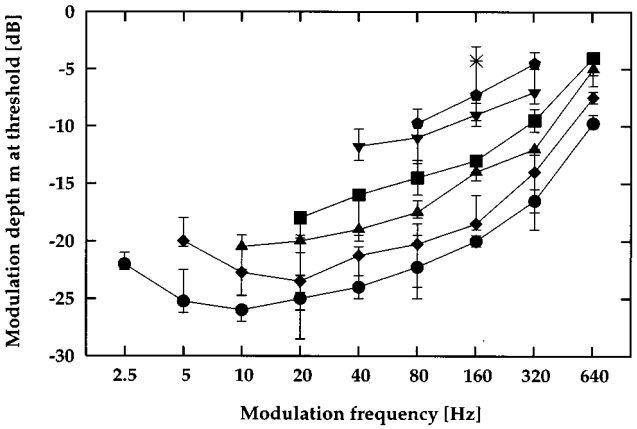


FIG. 10. Simulated TMTFs obtained with the current model for the same conditions as in Fig. 9.

poral integration in amplitude modulation detection. Several ''critical'' experiments were performed or taken from the literature and model predictions were compared with experimental data. As an extension to the single-channel model proposed in Dau *et al.* (1997), in which the concept of the modulation filterbank was established and the parameters of the modulation filters were fitted to data for narrow-band carriers, the modulation analysis was applied to broadband conditions. In this multi-channel version of the model, the temporal envelopes of the stimuli were processed by the same modulation filterbank in parallel at the output of each stimulated peripheral filter and the decision device combined all filter outputs linearly.

A. Spectral integration

In the modulation-detection conditions considered in this study using stochastic noise carriers and signal modulation, which is coherent across auditory filters, peripheral filtering generally does not limit the ability to detect modulation. This has been interpreted in some former studies as the observers' ability to increase their effective bandwidth in detecting wideband signals (Green, 1960; Bos and de Boer, 1966; Schacknow and Raab, 1976). As already discussed by Viemeister (1979), the mechanism behind such a combination and the stage of processing where it occurs are still unclear. It has been suggested in the literature (Green and Swets, 1966) that in detection experiments with multicomponent signals, combination probably occurs at a high level: The observer can combine information nearly optimally from widely spaced critical bands. This model approach was adopted within the current study and may be denoted as an ''auditory-filter-based approach.'' In contrast, in models for modulation-detection conditions, it has been previously assumed that the combination of peripheral filter outputs occurs at a very early stage of processing (Viemeister, 1979; Berg, 1996). In this context, the model proposed by Viemeister (1979) that includes a wide predetection bandpass filter ($\Delta f = 2000$ Hz) followed by a nonlinearity and a low-pass filter, can account for modulation detection data using (broadband) noise as the carrier. This model approach may be denoted as ''predetection-filter approach.'' For low modulation frequencies and broadband carriers, it is difficult to discriminate between such a predetection-filter approach and the auditory-filter-based approach.

Furthermore, it is not possible to discriminate whether there is one ''large'' modulation filterbank behind the combined outputs of *all* peripheral channels (cf. Yost *et al.*, 1989) or, alternatively, whether there is a modulation filterbank that separately analyzes the output of each peripheral channel before the information is combined. An argument in favor of the auditory-filter-based approach is that the bandwidth of the predetection filter of the Viemeister model is larger than the critical band estimates for most of the auditory range. This latter approach therefore fails to describe the data in an appropriate way for conditions in which spectral resolution of the auditory system plays any role. The present model gives a more general description of the processing of modulation in the auditory system and is also applicable to spectral-masking conditions, as was shown in Verhey and Dau (1996).

One of the main results of this study is that all the data could be described in terms of masking phenomena in the modulation-frequency domain. Analogous to the results described in Dau *et al.* (1997), thresholds are mainly determined by the amount of the *inherent* modulation power of the specific noise carrier that falls into the transfer range of the modulation filter tuned to the test modulation frequency. The model can therefore account for the experimental findings of Eddins that time constants derived from the TMTF do not vary with changing frequency region and decrease with increasing stimulus bandwidth. Note, however, that this model does not support the notion of one resolution time constant derivable from the data because the low-pass characteristic in the modulation data for broadband noise is not caused by a corresponding ''low-pass weighting'' of fast envelope fluctuations by the auditory system. Instead, the data are explained in terms of the interaction of stimulus power in the modulation-frequency region and the scaling of the modulation filters.[6] We want to add that, recently, Strickland and Viemeister (1997) have pointed out that the effective change of time constant with carrier bandwidth in Eddins' experiment might be an artifact caused by the stimulus generation procedure. They showed experimentally that without the bandlimitation after filtering the dependence of the time constant on the carrier bandwidth is strongly reduced.

B. Temporal integration

The phenomenon of temporal integration refers to the fact that, over a range of durations, thresholds decrease with increasing signal duration. Several models have been suggested in the literature that describe the phenomenon of temporal integration. For example, to describe the threshold function observed in typical test-tone integration data, an integration process is assumed that occurs over relatively long time periods, i.e., of the order of several hundred milliseconds. The most prominent approach is the energy detection model (Green, 1960; Green and Swets, 1966) in which decisions are based on the power of the input integrated over a fixed time period. Another class of models assumes a shorter integration time to account for temporal resolution, such as modulation detection (Viemeister, 1979), gap detection (Forrest and Green, 1987), and temporal aspects of nonsimultaneous masking (Moore *et al.*, 1988; Oxenham and Moore, 1994). The discrepancy between these two modeling strategies is often described as illustrating the ''resolution-integration'' paradox (de Boer, 1985).

In recent studies, however, it has been argued that the disparity between the integration and the resolution time constants is not a real problem (Viemeister and Wakefield, 1991). They pointed out that the observation of a 3-dB decrease in threshold for each doubling of duration—as seen in typical test-tone integration data—means that the auditory system behaves *as if* perfect power integration occurs but that the system is not necessarily performing the operation of mathematical integration. Therefore it might be important to distinguish between the phenomenon of temporal integration and the process that accounts for the phenomenon. Viemeis-

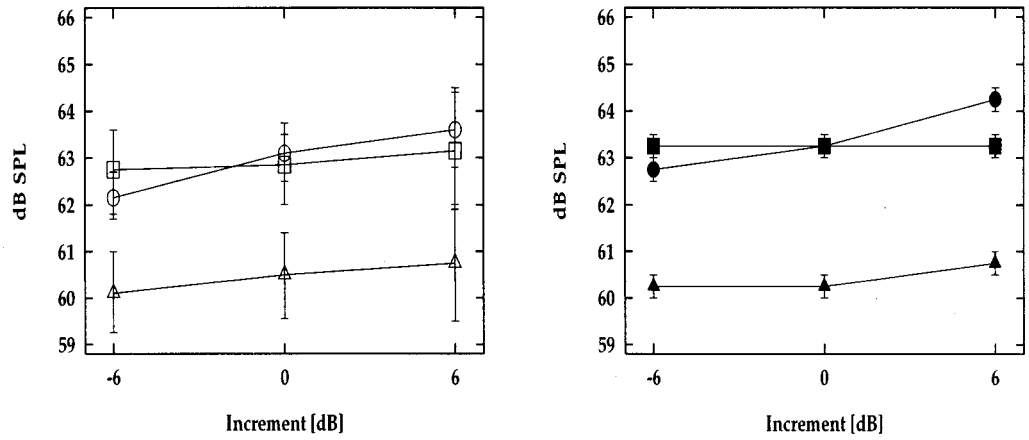


FIG. 11. Thresholds of a pair of tones separated by 100 ms as a function of the relative level of the intervening noise (cf. Viemeister and Wakefield, 1991). Thresholds for either the first or the second tone pulse alone are included for comparison. Squares: first pulse only; circles: second pulse only; triangles: pulse pair. Pulse(s): 10 ms, 1 kHz, presented during 10-ms gaps in a steady-state noise masker. During the 50-ms interval centered between the gaps, the noise level was either incremented by 6 dB, decremented by 6 dB, or left unchanged. Noise level: 40-dB spectrum level (measured at 1 kHz). Left panel: experimental data from Viemeister and Wakefield (1991). Right panel: corresponding simulated thresholds on the basis of the present model.

ter and Wakefield (1991) provided evidence that integration with a long time constant, such as proposed by the classical models, does not occur in all situations of auditory processing. They showed that the threshold for a pair of short pulses yields classic power integration only for pulse separations of less than 5–10 ms. For longer separations, the thresholds did not change with separation and the pulses appeared to be processed independently (cf. Zwislocki *et al.*, 1962). In a second experiment, Viemeister and Wakefield (1991) showed that the threshold for a pair of tone pulses was lower than for a single pulse, indicating some type of integration, but was not affected by level changes of the noise which was presented between the two pulses. The experimental results from that study are plotted in the left panel of Fig. 11. It shows the average thresholds for the first pulse alone (squares), the second pulse alone (circles), and for the pulse pair (triangles) as a function of the relative level of the intervening noise. The thresholds for the first pulse alone do not depend on the noise level. There is a slight increase in threshold for the second pulse reflecting forward masking. The thresholds for the pulse pair are about 2.5 dB lower than those for either pulse alone and do not depend on the level of the intervening noise (for details see Viemeister and Wakefield, 1991). These data cannot be explained by long-term integration.

Furthermore, as discussed by Viemeister and Wakefield (1991), the results of this second experiment are also inconsistent with the model proposed by Penner (1978). Penner showed that a compressive nonlinearity followed by a short-time constant integration can result in long *effective* integration. However, such a model—as an example of the class of ''single-look'' integration models—would predict a certain change in threshold depending on the energy of the noise between the two pulses, since the lower threshold for a pulse pair compared to that for the single pulse requires integration at least over the time of separation of the two pulses.

To account for the data, Viemeister and Wakefield (1991) proposed a ''multiple-look'' model. With such a model, ''looks'' or samples from a short-time constant process are stored in memory and can be processed ''selectively,'' depending on the task. The short-time constant allows the model, in principle, to account for temporal resolution data. The combination of a short-time constant and selective processing allows the model to also account for the data from the pulse pair experiment described above. The authors suggested that in temporal integration tasks, the long effective time constants may result from the combination of information from different looks, not from true long-term integration. However, there are some open questions with regard to the multiple-look model proposed by Viemeister and Wakefield (1991). A very basic question is concerned with what is meant by a ''look.'' The authors discuss the question of whether a look may be considered as a sample from the envelopelike waveform from the short temporal window that defines a look. But what is the shape of such a window? What about the correlation between successive looks or samples and how are these looks or samples combined to arrive at a decision statistics? The predicted temporal integration function described in the study of Viemeister and Wakefield (1991) depends strongly on how the looks are weighted and combined.

Nevertheless, the basic concept of the multiple-looks approach is to take into account that the observer attempts to use *all* the samples from the observation interval. For the detection of a tone, for example, an increase in the duration of the tone increases the number of samples and results in an improvement in performance. The model proposed in the present study contains an optimal detector as a decision device. The detection process can be considered as a ''matched filtering'' process as already described in Dau *et al.* (1996a). This implies that a *variable* time constant is available that is matched to the signal duration, dependent on the specific task, i.e., the model has at its disposal a continuum of time constants. The integration of the cross correlator is similar to the classic notion of temporal integration, but no fixed integration time constant is necessary for long-term integration.

It is the temporal extension of the template which automatically determines the weighting of stimuli across time. Thus our implementation is effectively close to the ''multiple-look'' strategy discussed by Viemeister and Wakefield (1991). Time constants that are related to the ''hard-wired'' part of signal processing within the model represent a lower limit of temporal acuity. (The term ''hard-wired'' is used in the sense that this part of signal processing is assumed to be independent of the specific experimental task.) The modulation filterbank represents a set of time constants that are, however, too short to account for the long-term integration data. Thus it is the decision device that inherently accounts for the long effective time constants. This agrees well with the considerations by Viemeister and Wakefield (1991) that different strategies are probably being employed, and different capabilities tapped, in resolution and integration tasks. The resolution task seems to use more ''peripheral'' processes whereas temporal integration may require higher level processes such as multiple sampling and probability summation. To that extent, assuming that the decision process occurs at a higher level of auditory processing, there is a certain correspondence between the modeling strategy of the present model and that suggested in the ''multiple-look'' model. The current approach may therefore be considered as an alternative strategy to the multiple-look model.

This correspondence is supported by the model predictions for the experiment of Viemeister and Wakefield, which are shown in the right panel of Fig. 11. Our model predicts a 3-dB decrease in threshold for the pulse pair compared to the threshold for a single pulse. Such a result is expected if the psychometric function for detection of pure tones in noise and in quiet can be described by: $d' = m(E/N_0)^k$, where d' is the normalized sensitivity index, E is signal power, N_0 is the spectral noise power density, and k and m represent ''individual'' parameters. For *ideal* observers, the parameter k has the value 0.5 (cf. Green and Swets, 1966), leading to a 3-dB decrease in threshold for the pulse pair compared to that for a single pulse. According to Egan *et al.* (1969), the psychometric function for normal hearing subjects can be described more accurately with a value $k \approx 1$ instead of 0.5. Such a value would lead to a 1.5-dB instead of 3-dB lower threshold for the pulse pair than for a single pulse, as was discussed by Viemeister and Wakefield (1991).

Long effective time constants occur both in typical test-tone integration and in the modulation integration examined in the present study. Whereas the decision device is responsible for the *shift* of the TMTF with changing signal duration, from which an *integration time constant* can be derived, it is the scaling of the modulation filterbank that determines the *form* of the TMTF, from which a *resolution time constant* is commonly derived. As discussed in the section about spectral integration in modulation detection, the threshold is determined mainly by the portion of the modulation power of the broadband noise carrier that is processed by the modulation filters tuned to the signal modulation frequency.

C. Future extensions of the model

Generally, all predicted thresholds shown in this paper lie between 2 and 5 dB lower than the experimental data. This shift indicates that there must be a certain loss of information in the auditory processing of modulation (independent of modulation rate) that is not at present accounted for in the model. Further modeling efforts are required to understand this discrepancy, that cannot be explained by simply increasing the variance of the internal noise, since thresholds are mainly determined by the external statistics of the stimuli.[7]

The present model does not cover conditions which require some processes of *across-channel comparison*. Such across-critical-band processing of temporally modulated complex stimuli might occur in, for instance, comodulation masking release (CMR) (Hall, 1987), comodulation detection difference (CDD) (McFadden, 1987; Cohen and Schubert, 1987), and modulation detection interference (MDI) (Yost and Sheft, 1989). In these cases the auditory system seems to be looking across frequency channels that contain temporally modulated stimuli. Concerning conditions of MDI, Yost *et al.* (1989) suggested a ''large'' modulation filterbank in which information about modulations of comparable rates are combined across the whole frequency range. Such an approach appears interesting. However, to simultaneously ensure the possibility of predicting *spectral* masking, it appears necessary to process a certain low-frequency part of the modulation spectrum including the dc-component *separately* within each peripheral channel, because this component represents the energy. Higher rate modulation may be combined within one large modulation filterbank. This, of course, should be tested in further studies in this field.

In order to predict performance for these types of experiments (as well as binaural masking experiments), additional stages would have to be included that calculate, for instance, the correlation between the envelopes of different frequency regions (or the two ears). Such stages have not been systematically evaluated within the framework of this study (for a first result on binaural masking, see Holube *et al.*, 1995a, b). However, using the present model as a preprocessing circuit, it might be easier to find realistic across-channel processing stages that also allow correct prediction for conditions in which across-channel processing is *not* needed.

V. CONCLUSIONS

The multi-channel modulation filterbank model described in this study can predict a wide variety of experimental conditions, including spectral and temporal integration of modulation detection and modulation masking with broadband carriers.

(1) Spectral integration is accounted for by combining the detection cues from all auditory filters with an optimal decision statistic.

(2) Temporal integration is accounted for by the variable length of the template that forms the basis of the optimal detector incorporated in the model.

(3) The combination of the modulation filterbank concept and the optimal decision algorithm presents a powerful strategy for describing modulation detection phenomena in narrow-band and broadband conditions.

ACKNOWLEDGMENTS

We would like to thank all our colleagues of the Graduiertenkolleg Psychoakustik at the University of Oldenburg for fruitful discussions on the content of this paper. We also thank Brian Moore, Jesko Verhey, Andrew Oxenham, Stefan Münkner, and Ralf Fassel for their comments and suggestions concerning this study and for their critical reading of earlier versions of this paper. The two anonymous reviewers also provided very constructive criticism.

[1]The implementation described in this paper represents one of two slightly different modulation filterbank models that were developed and tested in parallel at the universities of Oldenburg and Göttingen. While this paper concentrates on conditions relevant in the context of modulation perception, the paper by Münkner and Kohlrausch (1997) focuses on other measures of temporal processing such as gap detection and forward masking.

[2]This statement from Maiwald (1967b) is in contrast to some data shown in his thesis (Maiwald, 1966, Fig. 26). This figure shows that modulation-detection thresholds obtained with a noise carrier ranging from 0 to 16 kHz are about 3 dB lower than those for a carrier ranging from 6.5 to 16 kHz [modulation-detection thresholds expressed as $20 \log(m)$].

[3]These results of Houtgast are in contrast with data from a later study by Grantham and Bacon (1991) who were not able to replicate Houtgast's bandwidening results, despite increasing the bandwidth by two octaves. The authors concluded from their study that the effects observed in modulation-masking experiments may be explained on the basis of some kind of temporal pattern discrimination and not on a critical modulation band filtering process. As the authors state, however, it is possible that different methods of producing the noise modulators and of combining noise and signal may account for the discrepancy between their results and those of Houtgast (1989).

[4]It is assumed that adaptation effects are responsible for this effect, reflecting the temporal relationship between the carrier onset and the onset of modulation. The carrier onset produces a change in the level of the internal representation that is large relative to the changes produced by the sinusoidal modulation and might therefore interfere with the detection of the modulation, especially at low modulation frequencies (Viemeister, 1979).

[5]Sheft and Yost (1990) found that even when presenting the carrier with a fixed duration that included a 500-ms carrier fringe *preceding* the onset of modulation, threshold patterns exhibit a bandpass characteristic. The adaptation model of Püschel (1988) would not account for such a long-term adaptation effect.

[6]If frozen-noise carriers were used simulated thresholds obtained with the present model would be mainly determined by the statistics of the *internal* noise (added after the nonlinear processing). This would lead to significantly lower thresholds than for running-noise carriers. Threshold differences between both types of carriers observed in our own experimental data appear to depend on carrier bandwidth, with a large effect at very small bandwidths and a decreasing difference with increasing bandwidth. Studies on capabilities and limitations of the present model in conditions with frozen-noise carriers are currently in progress.

[7]The amount of the variance of the internal noise was determined in a simulation on intensity discrimination with deterministic stimuli to satisfy Webers law. A simple increase of the amount of internal noise necessary to compensate the observed 2–5 dB discrepancy in the present study would lead to a completely unrealistic threshold prediction in the calibration condition.

Berg, B. G. (**1996**). ''On the relation between comodulation masking release and temporal modulation transfer functions,'' J. Acoust. Soc. Am. **100**, 1013–1024.

Boer, E. de (**1985**). ''Auditory Time Constants: A Paradox?'' in *Time Resolution in Auditory Systems*, edited by A. Michelsen (Springer-Verlag, Berlin), pp. 141–158.

Bos, C. E., and Boer, E. de (**1966**). ''Masking and discrimination,'' J. Acoust. Soc. Am. **39**, 708–715.

Cohen, M. F., and Schubert, E. T. (**1987**). ''The effect of cross-spectrum correlation on the detectability of a noise band,'' J. Acoust. Soc. Am. **81**, 721–723.

Dau, T. (**1996**). ''Modeling auditory processing of amplitude modulation,'' Doctoral thesis, University of Oldenburg.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996a**). ''A quantitative model of the ''effective'' signal processing in the auditory system: I. Model structure,'' J. Acoust. Soc. Am. **99**, 3615–3622.

Dau, T., Kollmeier, D., and Kohlrausch, A. (**1997**). ''Modeling auditory processing of amplitude modulation: I. Detection and masking with narrowband carriers,'' J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996b**). ''A quantitative model of the ''effective'' signal processing in the auditory system: II. Simulations and measurements,'' J. Acoust. Soc. Am. **99**, 3623–3631.

Eddins, D. (**1993**). ''Amplitude modulation detection of narrow-band noise: Effects of absolute bandwidth and frequency region,'' J. Acoust. Soc. Am. **93**, 470–479.

Egan, J. P., Lindner, W. A., and McFadden, D. (**1969**). ''Masking-level differences and the form of the psychometric function,'' Percept. Psychophys. **6**, 209–215.

Formby, C., and Muir, K. (**1988**). ''Modulation and gap detection for broadband and filtered noise signals,'' J. Acoust. Soc. Am. **84**, 545–550.

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.

Grantham, D. W., and Bacon, S. P. (**1991**). ''Binaural modulation masking,'' J. Acoust. Soc. Am. **89**, 1340–1349.

Green, D. M. (**1960**). ''Auditory detection of a noise signal,'' J. Acoust. Soc. Am. **32**, 121–131.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hall, J. W., III (**1987**). ''Experiments on comodulation masking release,'' in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ).

Holube, I., Colburn, H. S., van de Par, S., and Kohlrausch, A. (**1995a**). ''Model simulations of masked thresholds for tones in dichotic noise maskers,'' J. Acoust. Soc. Am. **97**, 3411–3412.

Holube, I., Colburn, H. S., van de Par, S., and Kohlrausch, A. (**1995b**). ''Simulationen der Mithörschwellen von Testtönen in dichotischen Rauschmaskierern,'' Fortschritte der Akustik, DAGA '95, pp. 783–786.

Houtgast, T. (**1989**). ''Frequency selectivity in amplitude-modulation detection,'' J. Acoust. Soc. Am. **85**, 1676–1680.

Lawson, J. L., and Uhlenbeck, G. E. (**1950**). *Threshold Signals*, Volume 24 of *Radiation Laboratory Series* (McGraw-Hill, New York).

Maiwald, D. (**1966**). ''Zusammenhang zwischen Mithörschwellen und Modulationsschwellen,'' Doctoral thesis, Technical University of München.

Maiwald, D. (**1967a**). ''Ein Funktionsschema des Gehörs zur Beschreibung der Erkennbarkeit kleiner Frequenz- und Amplitudenänderungen,'' Acustica **18**, 81–92.

Maiwald, D. (**1967b**). ''Die Berechnung von Modulationsschwellen mit Hilfe eines Funktionsschemas,'' Acustica **18**, 193–207.

McFadden, D. (**1987**). ''Comodulation detection differences using noise-band signals,'' J. Acoust. Soc. Am. **81**, 1519–1527.

Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). ''The shape of the ear's temporal window,'' J. Acoust. Soc. Am. **83**, 1102–1116.

Münkner, S., and Kohlrausch, A. (**1997**). ''Simulations of temporal resolution and integration experiments using a modulation filterbank model'' (in preparation).

Oxenham, A. J., and Moore, B. C. J. (**1994**). ''Modeling the additivity of nonsimultaneous masking,'' Hearing Res. **80**, 105–118.

Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1987**). ''An efficient auditory filterbank based on the gammatone function,'' in Paper presented at a meeting of the IOC Speech Group on Auditory Modelling at RSRE, 14–15 December.

Penner, M. J. (**1978**). ''A power law transformation resulting in a class of short-term integrators that produce time-intensity trades for noise bursts,'' J. Acoust. Soc. Am. **63**, 195–201.

Püschel, D. (**1988**). ''Prinzipien der zeitlichen Analyse beim Hören,'' Doctoral thesis, University of Göttingen.

Rodenburg, M. (**1972**). ''Sensitivity of the auditory system to differences in intensity,'' Doctoral thesis, Medical Faculty of Rotterdam.

Rodenburg, M. (**1977**). ''Investigations of temporal effects with amplitude modulated signals,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 429–437.

Schacknow, P. N., and Raab, D. H. (**1976**). ''Noise-intensity discrimination:

Effects of bandwidth conditions and mode of masker presentation,'' J. Acoust. Soc. Am. **60**, 893–905.
Schreiner, C., and Langner, G. (**1988**). ''Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanism,'' J. Neurophysiol. **60**, 1799–1822.
Sheft, S., and Yost, W. (**1990**). ''Temporal integration in amplitude modulation detection,'' J. Acoust. Soc. Am. **88**, 796–805.
Strickland, E. A., and Viemeister, N. F. (**1997**). ''The effects of frequency region and bandwidth on the temporal modulation transfer function,'' J. Acoust. Soc. Am. **102**, 1799–1810.
van Zanten, G. A. (**1980**). ''Temporal modulation transfer functions for intensity modulated noise bands,'' in *Psychophysical and Behavioral Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft U.P., Delft, The Netherlands), pp. 206–209.
Verhey, J. L., and Dau, T. (**1996**). ''Simulations of spectral masking with a model incorporating an optimal decision strategy,'' in *Psychoacoustics, Speech, and Hearing Aids*, edited by B. Kollmeier (World Scientific, Singapore), pp. 29–34.
Viemeister, N. F. (**1977**). ''Temporal factors in audition: a system analysis approach,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 419–427.
Viemeister, N. F. (**1979**). ''Temporal modulation transfer functions based upon modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380.
Viemeister, N. F., and Wakefield, G. H. (**1991**). ''Temporal integration and multiple looks,'' J. Acoust. Soc. Am. **90**, 858–865.
Yost, W. A., and Sheft, S. (**1989**). ''Across critical band processing of amplitude modulated tones,'' J. Acoust. Soc. Am. **85**, 848–857.
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.
Zwislocki, J. J., Hellman, R. P., and Verillo, R. T. (**1962**). ''Threshold of audibility for short pulses,'' J. Acoust. Soc. Am. **34**, 1648–1652.

Detection and discrimination of frequency glides as a function of direction, duration, frequency span, and center frequency

John P. Madden and Kevin M. Fire
Department of Communication Sciences and Disorders, University of North Dakota, P.O. Box 8040, University Station, Grand Forks, North Dakota 58202

(Received 30 January 1997; revised 21 July 1997; accepted 21 July 1997)

The study investigated the ability to detect and discriminate frequency glides under a variety of experimental conditions. The subjects distinguished between a comparison signal that either was level in frequency or was swept across a fixed frequency span, and a target signal that changed more in frequency than the comparison signal. Tone durations were 50 and 400 ms. Nominal center frequencies were 0.5, 2, and 6 kHz; actual center frequencies were varied randomly, or roved, over a range equal to 0.1 times the nominal center frequency. Up- and down-glides were used. The transition span of the comparison signal was either 0, 0.5, 1, or 2 times the equivalent rectangular bandwidth of the auditory filter at the nominal center frequency. Discrimination thresholds were obtained for all combinations of center frequency, direction, and span. Overall, thresholds expressed as ΔHz/ERB varied little as a function of center frequency. Glide duration had no effect on discrimination. The 50-ms down-glides were more difficult to detect than the 50-ms up-glides; otherwise, the effect of direction was not significant. With the exception of the 50-ms down-glides, detection/discrimination thresholds were similar for the 0-, 0.5-, and 1-ERB transition spans, but increased significantly for the 2-ERB span. The absence of significant variation across frequency supports a place mechanism for the detection of frequency change in gliding tones, based on the detection of changes in the excitation pattern. An excitation pattern model cannot account for the asymmetry noted for glide detection, however.
[S0001-4966(97)02311-4]

PACS numbers: 43.66.Ba, 43.66.Fe [WJ]

INTRODUCTION

Recently, Madden and Fire measured detection and discrimination thresholds for frequency glides (Madden and Fire, 1996). When expressed as a proportion of the equivalent rectangular bandwidth (ERB) of the auditory filter, thresholds varied little across center frequencies ranging from 0.5 to 6 kHz. This lack of significant variation in threshold across center frequency was interpreted as evidence for a place mechanism for coding frequency change that monitors the excitation pattern produced by the stimulus. Signal duration was fixed at 50 ms, however, and thus the study provided evidence supporting a place mechanism for signals of brief duration only. A ''sluggish'' temporal mechanism based on phase-locking might be used to track glides of greater duration and more gradual rate of change, as suggested by Sek and Moore (1995). Also, all signals were up-glides, and therefore the effect of direction was not investigated, nor has it been systematically investigated for glide discrimination in previous studies. Thus, the present study was undertaken to extend the initial study to determine the effect of duration and direction on glide detection and discrimination thresholds.

For the most part, previous studies have found that glide detection and discrimination thresholds decrease as signal duration increases, with most of the decrease having occurred when signal duration reaches 100 ms (Sergeant and Harris, 1962; Pollack, 1968; Tsumura *et al.*, 1973; Dooley and Moore, 1988a). However, in a glide detection task, if the center frequencies of the tones being distinguished are the same, the subject potentially can compare the end points of the tones and select the tone with the end point that is higher or lower in pitch. If this is the case, the subject is performing a between-signal pitch comparison which is similar to a DLF determination, rather than estimating the pitch change within each glide. It is possible that the subjects in these studies were making such end-point comparisons. The utility of this cue could increase as signal duration increases because the subjects could obtain a longer-duration sample at the end point of the signal. In the present study, center frequencies were roved over a range of frequencies that was a fixed proportion of the nominal center frequency, to force the subjects to use only within-signal information.

Several studies have noted direction asymmetries for glide detection, but the results are contradictory. Dooley and Moore (1988a) found that at most durations and center frequencies tested, average glide detection thresholds for up-glides were larger than those for down-glides, although the differences were small except at 2 kHz. Tsumura *et al.* (1973) found an asymmetry in the same direction and of a similar magnitude, with the up-glide thresholds being about 20% greater than the down-glide thresholds. Arlinger *et al.* (1977) and Nabelek and Hirsh (1969) found no asymmetry. Finally, both Gardner and Wilson (1979) and Schouten (1985) found that detection thresholds for down-glides were greater than those for up-glides. The present study in-

vestigated the effect of glide direction in greater depth than these previous studies by measuring thresholds for up- and down-glides across a range of center frequencies in several glide discrimination conditions, as well as a glide detection condition.

The overall purpose of the study was to investigate further the mechanisms that underlie glide perception. The primary objective was to determine whether there is evidence for temporal coding of long-duration glides, by comparing detection/discrimination thresholds for 50- and 400-ms glides with fixed frequency spans. Based on the results from the authors' previous experiments (Madden and Fire, 1996), it was expected that thresholds for the 50-ms signals would be a constant proportion of ERB across center frequency in both the up and down conditions at all transition spans, thus supporting a place mechanism. In contrast, it was hypothesized that a temporal mechanism based on phase-locking to the fine structure of the signal might be involved in the detection and discrimination of the more slowly changing 400-ms signals. Evidence for this would be a shift from the ''constant proportion of ERB'' threshold pattern for the 50-ms signals to one that is similar to the pattern of the DLF for pulsed tones in the case of the 400-ms signals. The latter pattern generally is thought to be determined by the operation of a temporal mechanism in the low frequencies and a place mechanism in the high frequencies (Moore, 1973). DLF thresholds typically are relatively constant from 0.5 to 2 kHz, but increase by a factor of 3 or 4 at 6 kHz (see, e.g., Moore and Glasberg, 1989). It was thought that this pattern of threshold variation across frequency would be most evident for the glide detection conditions, in which the comparison signal does not change and the target signal changes very little, giving the ''sluggish'' temporal mechanism the greatest opportunity to operate. This pattern should gradually become less evident in the discrimination conditions as the frequency span of the standard signal increases, and with it the rate of change of the signals. A secondary objective was to determine whether direction affects glide detection and discrimination. Neither a place nor a temporal mechanism of frequency coding predicts that there should be an effect of direction.

I. METHOD

A. Subjects

Four subjects with normal auditory sensitivity participated in the experiment. All had hearing thresholds of 15 dB HL or less at the audiometric test frequencies and were paid for their participation.

B. Procedure

In a two-alternative, forced-choice task, the subjects were asked to identify a target signal which either increased or decreased in pitch (for the glide detection condition) or increased or decreased more in pitch than the comparison signal (for the glide discrimination conditions). Center frequencies were 0.5, 2, and 6 kHz, and the start and end points of the stimuli were linearly equidistant from their center frequencies. Stimulus durations were 50 and 400 ms, with additional rise–fall times of 5 ms. A cosine-squared windowing function was used. Thresholds were determined for each of the four combinations of duration and direction using comparison signals that changed in frequency by the following amounts: 0, 0.5, 1, and 2 ERB. Table I gives the transition spans used at the various center frequencies for the three gliding comparison signals. All signals were presented monotically at 70 dB SPL, with an interstimulus interval of 500 ms.

TABLE I. Transition spans of the comparison signals, expressed in Hz. ERB values were calculated using the formula $\mathrm{ERB} = 24.7(4.37F + 1)$ from Glasberg and Moore (1990).

	Transition span		
Frequency	0.5 ERB	1 ERB	2 ERBs
500	40	79	158
2000	121	241	482
6000	336	672	1344

The center frequencies of the signals were roved over a range equal to 0.1 times the nominal center frequency. Roved center frequencies were randomly drawn from a uniform distribution. The center frequencies of both the comparison signal and the target were roved independently both within and between trials. A more detailed description of this procedure can be found in Madden and Fire (1996).

Thresholds were measured using an adaptive procedure that estimated the 70.7% correct response point on the psychometric function (Levitt, 1971). After two consecutive correct responses, the change in frequency of the target signal was decreased; after one incorrect response it was increased. Step size was decreased after the first six reversals and varied depending on the center frequency and condition. At a center frequency of 0.5 kHz, for the 0-Hz condition, the initial step size was 10 Hz, at the higher center frequencies and for the 1-and 2-ERB conditions, where thresholds were greater, initial step sizes ranged up to 40 Hz. Final step sizes ranged in a similar fashion between 2 and 10 Hz. The frequency increase of the target signal was not permitted to fall below that of the comparison signal.

A run consisted of 80 trials, with a break after the first 40, to maintain alertness. The first six reversals were discarded, and threshold was computed as the arithmetic mean of the subsequent reversals (always at least eight). The subjects were practiced for at least 16 hours before data collection began. For each combination of center frequency, duration, and transition span, they then were tested until performance did not consistently improve over six runs. Of these six, the runs that varied by more than one standard deviation from the mean were discarded, and the threshold calculated from the remaining data (always at least four runs). Data collection was completed for each center frequency-duration-direction combination before the next combination of conditions was begun, to facilitate optimal learning of detection/discrimination cues. The order of the duration, direction, and center frequency conditions was counterbalanced across subjects.

Stimulus presentation and response acquisition were

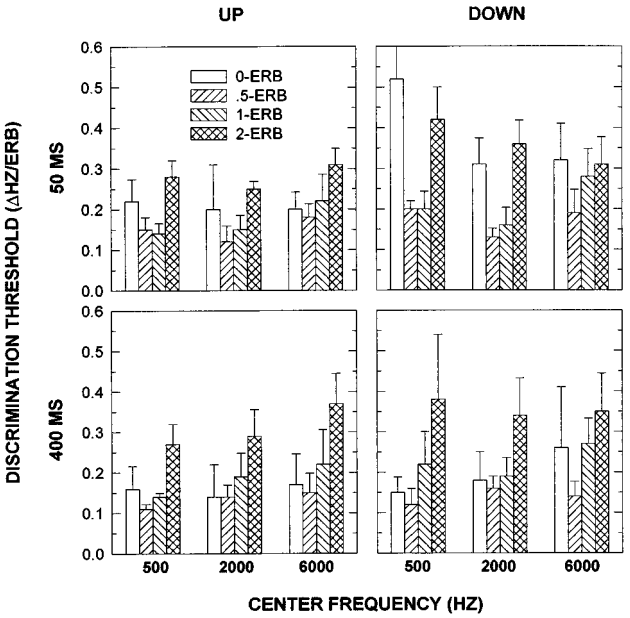


FIG. 1. Average discrimination thresholds for the various combinations of duration, direction, and frequency span, plotted as a function of center frequency. Thresholds are expressed as a proportion of the ERB at the signal center frequency. Error bars indicate one standard deviation from the mean.

controlled by a PC. Stimulus and response intervals were signaled on a computer monitor. The subjects received visual feedback after every trial, indicating the stimulus interval containing the target signal.

C. Stimulus generation

Digital representations of the stimuli were generated by an array processor (TDT-QAP2) running on a PC. A 16-bit-D-to-A converter (TDT-DD1) operating at a 50-kHz sampling rate converted the digital representations to analog waveforms which then were low-pass filtered at 8 kHz (TDT-FLT3), attenuated (TDT-PA3), and sent through an output amplifier (TDT-HBUF3) to an earphone (Sennheiser HD-414SL). The frequency response of this earphone has been reported to be relatively smooth across the range of the test frequencies (see Moore and Sek, 1995).[1] Signal spectra were verified using Hypersignal/Workstation spectrum analysis software.

II. RESULTS

Inspection of the data suggested that whereas absolute performance differed among the subjects, all demonstrated highly similar patterns of performance across the various experimental conditions. This supposition was tested statistically with a series of two-way ANOVAs (analyses of variance) which showed that the subject factor did not interact with the factors of duration, direction, center frequency, or transition span. Therefore, the data were collapsed across subject, and the average thresholds for the various experimental conditions are presented in Fig. 1. Thresholds are expressed as a proportion of the ERB at the nominal center frequency of the signal.

An inspection of the four graphs in Fig. 1 reveals that overall there is a tendency for thresholds to be lower at 2

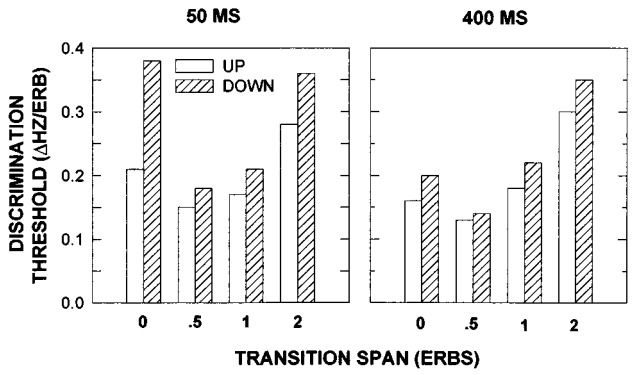


FIG. 2. Detection and discrimination thresholds for each signal duration, averaged across center frequency and plotted as a function of transition span.

than at 0.5 and 6 kHz, but that the differences in threshold magnitude between the center frequencies are not great. A one-way ANOVA indicated that there was not a significant main effect of center frequency [$F(1,176)=2.012$, $p=0.137$]. In addition, two-way ANOVAs with frequency and direction, frequency and duration, and frequency and transition span as factors revealed no significant interactions. Therefore, to get a clearer picture of the effects of direction, duration, and transition span, the data were collapsed across center frequency. The averaged thresholds are presented in Fig. 2.

It is evident from Fig. 2 that the 50- and 400-ms thresholds are very similar when compared within the same transition span/direction condition, except for the 0-ERB downglides. In this case, the 50-ms threshold is 1.9 times as great as the 400-ms threshold. Down-glide thresholds are greater than up-glide thresholds in all cases, but the differences are small, except for the 50-ms, 0-ERB thresholds. Here, the down-glide threshold is 1.8 times greater than the up-glide threshold. With the obvious exception of the 0-ERB, 50-ms, down-glide condition, thresholds are similar for the 0- to 0.5-, and 1-ERB transition spans, but increase considerably for the 2-ERB transition.

To test these conclusions statistically, a three-way ANOVA was conducted with duration, direction, and transition span as factors. There were significant main effects of duration [$F(1,176)=6.19$, $p=0.014$], direction [$F(1,176)=23.23$, $p<0.001$], and transition [$F(1,176)=38.338$, $p<0.001$]. There also were significant interactions between duration and direction [$F(3,176)=4.02$, $p=0.046$], direction and transition [$F(3,176)=2.63$, $p=0.05$], and duration and transition [$F(3,176)=6.37$, $p<0.001$].

To determine the sources of these effects, pairwise comparisons were conducted of the threshold values for all 16 combinations of the various levels of duration, direction, and transition span. The 50-ms, 0-ERB, down-glide threshold was significantly greater than the 50-ms, 0-ERB, up-glide threshold and the 400-ms, 0-ERB, down-glide threshold. There were no other significant differences associated with the effect of duration or direction. For the 400-ms signals, within both directions, the 0-, 0.5-, and 1-ERB thresholds did not differ from one another but differed from the 2-ERB thresholds. The same was essentially true for the 50-ms sig-

nals, with the exception of the 0-Hz, down-glide threshold, which differed from the corresponding 0.5- and 1-ERB thresholds, but did not differ from the corresponding 2-ERB threshold. Thus the statistical analysis supported the conclusions reached by inspection of the data.

Summarizing the main trends in the data, the effect of center frequency was not significant. There was no effect of duration or direction, except in the glide-detection condition, where there was an interaction between duration and direction. Here, the down-glide threshold was significantly greater than the up-glide threshold at 50 ms but not at 400 ms. Thresholds were similar for the 0-, 0.5-, and 1-ERB transition spans and increased for the 2-ERB span, with the exception of the 50-ms down-glides.

III. DISCUSSION

A. The effect of transition span

The overall average of the 2-ERB thresholds is slightly less than twice the overall average of the other transition spans, when the 50-ms, down-glide, 0-ERB threshold is not included. This was the pattern described for 50-ms up-glides by Madden and Fire in their 1996 paper, in which it was suggested that glide discrimination thresholds exhibit a constant Weber fraction.[2] The results of the current study extend this finding to include glides of long duration and down-glides, providing further support for the idea that glide discrimination thresholds follow Weber's law.

B. The effect of duration

The absence of a general effect of duration is surprising in view of the results from previous studies. In fact, the 400-ms, up-glide direction thresholds are considerably higher than those obtained by others using signals of similar duration. For example, Dooley and Moore (1988a) obtained average up-glide thresholds of about 4 Hz at 0.5 kHz and about 13 Hz at 2 kHz using 500-ms signals, whereas in the present experiment the corresponding 400-ms thresholds were 13 and 34 Hz. The most probable cause of this discrepancy is the use of signals roved in center frequency. Although a previous experiment (Madden and Fire, 1996) found that the effect of roving signals was not significant for the detection of 50-ms frequency glides, an ensuing pilot study indicated that the effect of roving the signals increases greatly when signal duration is increased. For example, an average detection threshold of 13 Hz was obtained for a 2000-Hz, 200-ms up-glide in the nonroved condition versus a roved threshold of 38 Hz for the same signal. It was speculated that the subjects used between-signal, end-point cues in the nonroved condition for these longer-duration signals, as was discussed in the Introduction. It thus appears that although glide detection thresholds decrease with increasing signal duration when center frequencies are constant, they do not change with duration (over the range tested here) when center frequencies are roved.

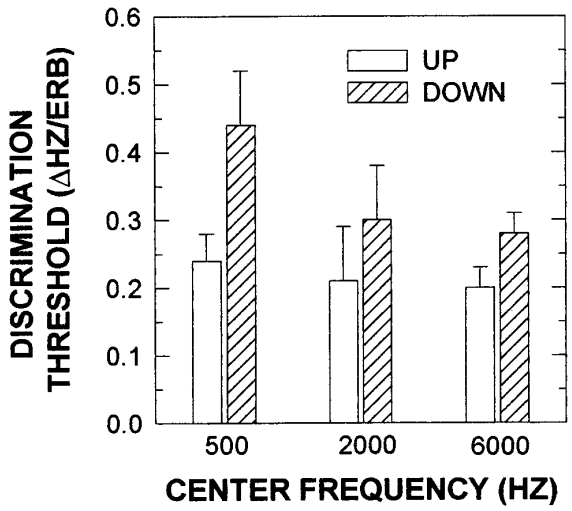


FIG. 3. Average thresholds for the 50-ms, 0-ERB condition from a second set of subjects, plotted as a function of center frequency. Error bars indicate one standard deviation from the mean.

C. The effect of direction

Because an effect of direction was noted only for the 50-ms, glide detection portion of the experiment, it was decided to recheck these thresholds by testing a second set of subjects consisting of two of the original subjects and two new subjects. The results are presented in Fig. 3 and are similar to those of the main experiment. A two-way ANOVA with frequency and direction as factors revealed significant main effects of direction $[F(1,18)=14.43,\ p=0.001]$ and center frequency $[F(2,18)=3.61,\ p=0.048]$ but no interaction $[F(2,18)=1.66,\ p=0.219]$. Pairwise comparisons indicated that the up and down thresholds differed significantly at 0.5 kHz, but not at 2 or 6 kHz. The greatest asymmetry was at 0.5 kHz in the main experiment, also. It was pointed out in the Introduction that other workers have reported results showing the opposite asymmetry, but including the present study, there now are at least three studies in which the same asymmetry has been observed for brief duration glides (also Gardner and Wilson, 1979 and Schouten, 1985). It is not clear what sort of underlying mechanism might be responsible for the asymmetry, or why its occurrence should be affected by duration and frequency.

D. Modeling glide detection and discrimination

The data yield little support for temporal coding of the 400-ms signals. As was suggested in the Introduction, if a temporal mechanism plays a role in tracking glides, one would expect an effect of center frequency, with an increase in threshold magnitude at 6 kHz relative to thresholds at the lower center frequencies, particularly in the glide detection conditions. Moore and Sek (1995) and Sek and Moore (1995), for example, observed that FM detection thresholds expressed as a proportion of ERB increased by a factor of almost 2 at a carrier frequency of 6 kHz, relative to the lower carrier frequencies, when the modulation rate was 2 Hz. In fact, the 400-ms thresholds are more consistent with the predictions of a model in which frequency changes are detected

by a mechanism that monitors changes in the level of excitation at the edges of the excitation pattern generated by a stimulus (Zwicker, 1970; Moore and Sek, 1994). For such a model, the accuracy with which the start and end points of a glide can be estimated will be determined by the steepness of the excitation pattern evoked by the signal, which in turn is determined by the bandwidth of the auditory filters centered at or just below the start and end frequencies. An excitation pattern model thus predicts that detection and discrimination thresholds should be a fixed proportion of auditory filter bandwidth across frequency. However, an excitation pattern model cannot account for the direction asymmetry that was observed for glide detection. The absence of a duration effect also is problematic. The 400-ms signals remain near their start and end-point frequencies for a longer time period than the 50-ms signals. This extra time should allow more accurate sampling by whatever mechanism is involved, and result in better discrimination at longer durations.

IV. CONCLUSIONS

(1) Glide discrimination was not affected by center frequency, duration, or direction.
(2) There was an effect of frequency transition span. Except for the 50-ms down-glides, thresholds remained relatively constant as the comparison signal span increased from 0 to 1 ERB, but almost doubled when the comparison signal span increased to 2 ERB.
(3) Glide detection thresholds were affected by an interaction between direction, duration, and transition span: 50-ms down-glides were more difficult to detect than 50-ms up-glides or 400-ms down-glides and up-glides. This effect was most evident at 0.5 kHz.
(4) The results are better accounted for by a place mechanism of frequency coding based on the detection of changes in the excitation pattern generated by the signal than by a temporal mechanism based on phase-locking to signal fine structure. However, the absence of a duration effect and the presence of a direction asymmetry are not explained by an excitation pattern model.

ACKNOWLEDGMENT

This research was supported by research Grant No. 1 R15 DC 02662-01 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

[1]The frequency response of the HD-414SL earphone exhibits the greatest amount of variation in the higher frequencies. A difference in change in level between signals to be discriminated is, of course, a potential discrimination cue, if it is large enough. For example, the average change in level for a 2-ERB, 6-kHz standard signal (transition span=1344 Hz) is about 4.4 dB. For a corresponding target signal with a transition span of 1744 Hz, which is about 100 Hz greater than the average target span at discrimination threshold for the 2-ERB condition in the previous study (Madden and Fire, 1996), the average change in level is about 5.7 dB. Thus the change in level of the target in this case would be about 1.3 dB greater than that of the standard. Dooley and Moore (1988b) found that subjects could detect a difference in change in level as small as 0.2 dB for 750-ms signals at 2 kHz. This would indicate that the detection of level changes is a potential cue for frequency glide detection and discrimination at the 6-kHz center frequency. However, the signals used by Dooley and Moore either did not change in frequency or changed by only a relatively small amount in comparison with the signals used in the glide discrimination tasks in the current study. Thus it is difficult to know the extent to which the findings of Dooley and Moore are relevant to the present study. Also, when the subjects in Madden and Fire (1996) were informally asked to comment on discrimination cues, they did not mention level changes and appeared unaware that any occurred.

[2]The authors are indebted to Brian C. J. Moore for suggesting this interpretation of the data reported in Madden and Fire (1996).

Arlinger, S. D., Jerlvall, L. B., Ahren, T., and Holmgren, E. C. (**1977**). ''Thresholds for linear frequency ramps of a continuous pure tone,'' Acta Oto-Laryngol. **83**, 317–327.

Dooley, G. J., and Moore, B. C. J. (**1988a**). ''Detection of linear frequency glides as a function of frequency and duration,'' J. Acoust. Soc. Am. **84**, 2045–2057.

Dooley, G. J., and Moore, B. C. J. (**1988b**). ''Duration discrimination of steady and gliding tones: A new method for estimating sensitivity to rate of change,'' J. Acoust. Soc. Am. **84**, 1332–1337.

Gardner, R. B., and Wilson, J. P. (**1979**). ''Evidence for direction-specific channels in the processing of frequency modulation,'' J. Acoust. Soc. Am. **66**, 704–709.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hearing Res. **47**, 103–138.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Madden, J. P., and Fire, K. M. (**1996**). ''Detection and discrimination of gliding tones as a function of frequency transition and center frequency,'' J. Acoust. Soc. Am. **100**, 3754–3760.

Moore, B. C. J. (**1973**). ''Frequency difference limens for short-duration tones,'' J. Acoust. Soc. Am. **54**, 610–619.

Moore, B. C. J., and Glasberg, B. R. (**1989**). ''Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation,'' J. Acoust. Soc. Am. **86**, 1722–1732.

Moore, B. C. J., and Sek, A. (**1994**). ''Effects of carrier frequency and background noise on the detection of mixed modulation,'' J. Acoust. Soc. Am. **96**, 741–751.

Moore, B. C. J., and Sek, A. (**1995**). ''Effects of carrier frequency, modulation rate, and modulation waveform on the detection of modulation and the discrimination of modulation type (amplitude modulation versus frequency modulation),'' J. Acoust. Soc. Am. **97**, 2468–2478.

Nabelek, I., and Hirsh, I. J. (**1969**). ''On the discrimination of frequency transitions,'' J. Acoust. Soc. Am. **45**, 1510–1519.

Pollack, I. (**1968**). ''Detection of rate of change of auditory frequency,'' J. Exp. Psychol. **77**, 535–541.

Schouten, M. E. H. (**1985**). ''Identification and discrimination of sweep tones,'' Percept. Psychophys. **37**, 369–376.

Sek, A., and Moore, B. C. J. (**1995**). ''Frequency discrimination as a function of frequency, measured in several ways,'' J. Acoust. Soc. Am. **97**, 2479–2486.

Sergeant, R. L., and Harris, J. D. (**1962**). ''Sensitivity to unidirectional frequency modulation,'' J. Acoust. Soc. Am. **34**, 1625–1628.

Tsumura, T., Sone, T., and Nimura, T. (**1973**). ''Auditory detection of frequency transition,'' J. Acoust. Soc. Am. **53**, 17–25.

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, The Netherlands).

Loudness of dynamic stimuli in acoustic and electric hearing[a)]

Chaoying Zhang and Fan-Gang Zeng[b)]
Auditory Perception Laboratory, Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 31 March 1997; revised 14 July 1997; accepted 14 July 1997)

Traditional loudness models have been based on the average energy and the critical band analysis of steady-state sounds. However, most environmental sounds, including speech, are dynamic stimuli, in which the average level [e.g., the root-mean-square (rms) level] does not account for the large temporal fluctuations. The question addressed here was whether two stimuli of the same rms level but different peak levels would produce an equal loudness sensation. A modern adaptive procedure was used to replicate two classic experiments demonstrating that the sensation of ''beats'' in a two- or three-tone complex resulted in a louder sensation [E. Zwicker and H. Fastl, *Psychoacoustics—Facts and Models* (Springer-Verlag, Berlin, 1990)]. Two additional experiments were conducted to study exclusively the effects of the temporal envelope on the loudness sensation of dynamic stimuli. Loudness balance was performed by normal-hearing listeners between a white noise and a sinusoidally amplitude-modulated noise in one experiment, and by cochlear implant listeners between two harmonic stimuli of the same magnitude spectra, but different phase spectra, in the other experiment. The results from both experiments showed that, for two stimuli of the same rms level, the stimulus with greater temporal fluctuations sometimes produced a significantly louder sensation, depending on the temporal frequency and overall stimulus level. In normal-hearing listeners, the louder sensation was produced for the amplitude-modulated stimuli with modulation frequencies lower than 400 Hz, and gradually disappeared above 400 Hz, resulting in a low-pass filtering characteristic which bore some similarity to the temporal modulation transfer function. The extent to which loudness was greater was a nonmonotonic function of level in acoustic hearing and a monotonically increasingly function in electric hearing. These results suggest that the loudness sensation of a dynamic stimulus is not limited to a 100-ms temporal integration process, and may be determined jointly by a compression process in the cochlea and an expansion process in the brain. A level-dependent compression scheme that may better restore normal loudness of dynamic stimuli in hearing aids and cochlear implants is proposed. [S0001-4966(97)05810-4]

PACS numbers: 43.66.Ba, 43.66.Cb, 43.66.Mk, 43.66.Nm [WJ]

INTRODUCTION

Models of loudness have been mostly based on steady-state sounds, and rarely validated using dynamic sounds with large temporal fluctuations. For pure tones, loudness (L) has been measured directly using the magnitude estimate technique and found to grow as a power function of sound intensity (I), $L = kI^{0.3}$ (Stevens, 1957). Fletcher (1940) established that the ear can be modeled as a bank of frequency filters with the filter bandwidth being termed ''critical bandwidth.'' Fletcher and Munson (1933, 1937) used the critical band concept implicitly in either an equal-loudness-contour-based empirical formula or a masking audiogram to calculate the loudness of complex sounds with broadband spectra. Based on the pioneering work by both Fletcher and Stevens, Zwicker formulated an excitation-pattern model to predict the loudness of complex sounds (e.g., Zwicker and Scharf, 1965). In Zwicker's model, the physical spectrum of a complex sound is first subject to a critical-band analysis, in which the ''critical-band intensity'' is calculated by adding spectral intensities within each critical band; then an ''excitation pattern'' of the complex sound is derived using the shape of masked thresholds; finally, the ''specific loudness'' is calculated using a power law transformation of the excitation level for each critical band and is added across critical bands to obtain the ''total loudness'' of the complex sound.

Zwicker and his colleagues used a two-tone complex to test the validity of the spectrally based loudness model. The loudness of a two-tone complex was found to be determined by adding the intensity of the two tones if they were within one critical band and by adding the loudness if they were in two widely separated critical bands (Zwicker *et al.*, 1957; Zwicker and Fastl, 1990). However, a small but consistent deviation from the spectrally based model was also noticed, in which the loudness of two tones exceeded the predicted value based on intensity addition when the frequency separation of the two tones was within 10 Hz and produced a sensation of ''beats.'' Under such conditions, the loudness of the two-tone complex appeared to be determined by the peak amplitude rather than the rms (root-mean-square) amplitude.

[a)]Parts of this paper were presented at the 19th Midwinter Research Meeting of the Association for Research in Otolaryngology, February 1996, St. Petersburg, Florida and the Third Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, December 1996, Honolulu, Hawaii.

[b)]Author to whom correspondence should be addressed. Electronic mail: zeng@hei.org

Zwicker *et al.* further explored this phenomenon using a three-tone complex in which the phase relationship of three tones was manipulated to produce different temporal waveform envelopes. Again, when the total bandwidth of the three-tone complex was less than 10 Hz, the waveform with greater peak amplitude (amplitude modulation) produced a louder sensation than the waveform with smaller peak amplitude (quasifrequency modulation). Perhaps because this deviation from the spectrally based model produced a relatively small effect and occurred at very low-frequency temporal fluctuations (less than 10 Hz or longer than 100 ms), the loudness of dynamic stimuli has been assumed to be due to the temporal integration process, with a time constant of about 75 ms (Zwicker and Fastl, 1990). Zwicker's model is now widely accepted and has formed the basis on which the international standard for calculating the loudness of various complex sounds is established (ISO532). Recently, Moore and Glasberg (1996) revised Zwicker's model to account more accurately for equal loudness contours and the loudness of partially masked sounds.

We are motivated to revisit the issue of loudness of dynamic stimuli based on questions arising from fitting prosthetic devices to hearing-impaired listeners. The first question relates to hearing aid gain prescription. To compensate for reduced audibility due to hearing loss, one desirable goal for a hearing aid is to amplify speech sounds appropriately in order to restore loudness to the most comfortable listening level. Most fitting procedures currently in use select the gain and frequency response from pure-tone thresholds and are based on empirical formulas (e.g., Byrne and Dillon, 1986). Although attempts have been made to use Zwicker's loudness model in hearing aid gain prescription, the results are not satisfactory because the model-predicted gain always overcompensates for the amount of hearing loss, and produces ''too loud'' sensations (Launer and Bachler, 1996). One reason for this failure is that Zwicker's loudness model calculates the rms level and does not account for temporal waveform fluctuations in dynamic stimuli such as speech sounds, which can have a peak amplitude of 10–20 dB above the rms level (Fletcher, 1953; Boothroyd *et al.*, 1994).

The second question relates to the loudness mapping problem in cochlear implants. In designing the implant processor, the amplitude of acoustic stimuli has to be compressed to accommodate the extremely narrow 10–20 dB dynamic range which is characteristic of the electric stimulation of the auditory nerve. Contrary to traditional compression circuits in hearing aids that have a fast attack time and a slow release time and are controlled by short-term averaging, the cochlear implant processor usually employs a zero attack- and release-time compression and employs a nonlinear compressive map between the instantaneous acoustic envelope magnitude and the electrical pulse amplitude (Wilson *et al.*, 1991). The rationale for compressing the instantaneous peak amplitude in cochlear implants was partially verified by Zeng and Shannon (1995), who found that the loudness of amplitude-modulated sinusoids was determined by the peak amplitude at an uncomfortable loudness level, but by the rms amplitude at the threshold level. The significant effect of the 100-Hz sinusoidal modulator on the loudness of the amplitude-modulated stimuli in Zeng and Shannon's study could not be explained simply on the basis of the presumably much slower temporal integration process.

The failure of applying Zwicker's loudness model to the fitting of prosthetic hearing devices can be due to many differences in analyzing sounds between normal-hearing and hearing-impaired listeners. One significant difference is the broader auditory filter in hearing-impaired listeners, or in the extreme case of single electrode stimulation of the auditory nerve, no spectral analysis mechanisms exist at all. Generally, the extent to which the temporal fluctuation of a broadband stimulus can be preserved is proportional to the ear's analysis (or critical) bandwidth. Compared with normal-hearing listeners, the temporal envelope of a dynamic stimulus is more likely to be preserved with broader auditory filters in impaired listeners and is preserved entirely with electric stimulation in implant listeners. The other significant difference is the steeper loudness growth in impaired listeners. Loudness recruitment is presumably due to a reduced level of compressive nonlinearity in cochlear-impaired listeners and a total loss in implant listeners. Several studies have attempted to modify Zwicker's model to take into account the factors of broader filters and loudness recruitment in hearing-impaired ears, but none of them was designed to deal with dynamic sounds (Florentine and Zwicker, 1979; Launer, 1995; Moore and Glasberg, 1997). There has been evidence that loudness recruitment not only occurs for steady-state sounds, but also accentuates the difference between peaks and valleys in dynamic stimuli (Moore *et al.*, 1996). If the temporal fluctuations can make a sound louder, then a greater temporal effect might be observed in hearing-impaired listeners.

The present study used two different methods to separate the effects of the critical band analysis from temporal processing in loudness sensation. First, we asked normal-hearing listeners to perform a loudness balance task between a white noise and a sinusoidally amplitude-modulated (SAM) noise. Both the white noise and the SAM noise had identical long-term spectra, and the only difference between them was the greater temporal fluctuations in the SAM noise. Thus if the temporal envelope had no effect on loudness sensation, then the white noise and the SAM noise would be equally loud as long as their rms levels were kept the same. Second, we asked cochlear implant listeners to judge the loudness of two harmonic stimuli with the same magnitude spectra but different phase spectra (Schroeder, 1970). Because manipulating phase relationships did not change the stimulus rms level, and the cochlear implant listeners were not able to resolve the harmonics at all, hypotheses regarding whether the rms level or the peak level determines loudness could be tested directly.

I. GENERAL METHODS

A. Subjects

Six normal-hearing listeners, including three males and three females, ranging from 25 to 35 years old, participated in this study. Four participated in the two- and three-tone experiments (exp. 1), after which two of them dropped out

and two additional listeners were recruited to participate in the SAM noise experiment (exp. 2). All subjects had normal hearing (10 dB HL or better pure-tone thresholds for frequencies between 125 and 5000 Hz). The subjects were seated in a double-walled acoustic booth and tested individually.

Three cochlear implant listeners participated in the study of balancing loudness between two harmonics of different phases (exp. 3). All cochlear implant listeners used the Ineraid device, which has a percutaneous plug that connects directly to intracochlear electrodes and allows undistorted electric stimulation with essentially any waveform. At the time of the test, the three implant listeners were 35–65 years old, and had used the device for more than seven years. All implant listeners had extensive experience in various psychophysical tasks. The most apical electrode was used as the stimulation electrode and the lead in the temporalis muscle was used as the ground electrode.

B. Stimuli

All stimuli were generated digitally by a PC using TDT (Tucker Davis Technologies) system II hardware. The stimuli were delivered through a 16-bit D/A converter at a 40-kHz sampling rate and smoothed by a 14-kHz low-pass filter. All stimuli had a 500-ms duration and 10-ms cosine-squared rise and fall times. In acoustic stimulation, the sound level was controlled digitally and calibrated periodically by a B&K sound level meter linked to a Zwislocki coupler and an ER-2 insert earphone (Etymotic Research). In electric stimulation, the output of the D/A converter was routed to a voltage-controlled current source that could deliver a maximal peak current of 1 mA (Vurek *et al.*, 1981). The speech processor of the cochlear implant was not used in the experiment. A foot-operated safety switch was provided between the current source and the percutaneous plug and could disconnect the circuitry immediately by either the experimenter or the listener in case of unpleasant or too loud stimulation.

C. Procedure

A two-interval, two-alternative, forced-choice, adaptive procedure was employed in which two randomly interleaved sequences with different decision rules bracketed the point of subjective equality in the loudness balance task (Jesteadt, 1980; Schlauch and Wier, 1987; Zeng and Turner, 1991). In these two sequences, the upper sequence used a 2-down, 1-up decision rule to track the 71% louder response level, while the lower sequence used a 1-down, 2-up decision rule to track the 29% louder level. In each trial, a standard sound with a fixed level and a comparison sound with a variable level that was selected randomly from one of two sequences were presented randomly in either the first interval or the second interval. Each interval was signaled by a light on the computer monitor and the two intervals were separated by 500 ms. The subject's task was to judge which one of the two intervals contained the louder sound by pressing one of two buttons on a PC mouse. No feedback regarding the ''correct'' response was given to the subject after each trial. Each sequence was terminated by either having a total of 75 trials or having reached 12 reversals, which corresponded to a change in the direction of louder response from the standard to the comparison stimulus or vice versa. The step size was 5–10 dB for the first four reversals and was reduced to 1 dB afterwards. The 10-dB step size was used initially in combination with a wide separation of the starting levels for the upper and lower sequences in order to locate quickly the point of subjective equality. The level at which a reversal occurred was recorded and the reversals, excluding the first four, were averaged at the end of the run to estimate the 71% and 29% louder levels. The point of subjective equality was estimated as the average of the 71% and the 29% levels. Each run took about 5–10 min and each data point reported in this study was estimated from 3–6 runs. The averaged standard deviation across all runs in all subjects was about 1.5 dB.

II. EXPERIMENT 1: LOUDNESS OF COMPLEX TONES IN ACOUSTIC HEARING

The purpose of this experiment was to use the modern adaptive procedure to replicate Zwicker's original two-tone and three-tone experiments. This adaptive procedure obtained relatively unbiased measures in the subjective task of balancing loudness between two sounds of different qualities (Jesteadt, 1980; Schlauch and Wier, 1987; Zeng and Turner, 1991). The adaptive procedure not only allowed us to estimate the mean magnitude of both the spectral analysis and the temporal envelope effects on loudness, but also allowed us to estimate the variability of these effects, because the 21% and the 79% levels approximated the plus and minus one-standard deviation from the mean in a 2-AFC task (Green, 1964).

A. Stimuli

The stimuli as described in the original two- and three-tone complex experiments (Zwicker and Fastl, 1990) were replicated in this experiment. For the two-tone complex standard, the two tone frequencies, f_1 and f_2, were geometrically centered on 1 kHz and had a frequency separation ($f_2 - f_1$) of 2, 5, 10, 20, 50, 100, 200, 500, 1000, and 2000 Hz. Each of the two tones was presented at a fixed level of 60 dB SPL, independent of the frequency separation. An example of the two-tone complex waveform is shown on the bottom panel of Fig. 1. Also shown is the waveform of the 1000-Hz comparison stimulus.

For the three-tone experiment, two types of standard stimuli were constructed to produce the same spectrum but different temporal envelopes. First, an amplitude-modulated (AM) stimulus was constructed with a modulation depth of 0.5 and an overall level of 45 dB SPL. The frequency spectrum of the AM stimulus had three components, including the carrier (f_c) and two side bands ($f_c - f_m$ and $f_c + f_m$, with f_m being the modulation frequency). Second, a quasi-frequency-modulation (QFM) stimulus was constructed identical to the AM stimulus, except that the carrier was 90 deg out of phase relative to the two side bands. Examples of the AM and QFM stimuli are shown on the bottom panel of Fig. 2. Notice the greater (1.9 dB) temporal envelope fluctuation

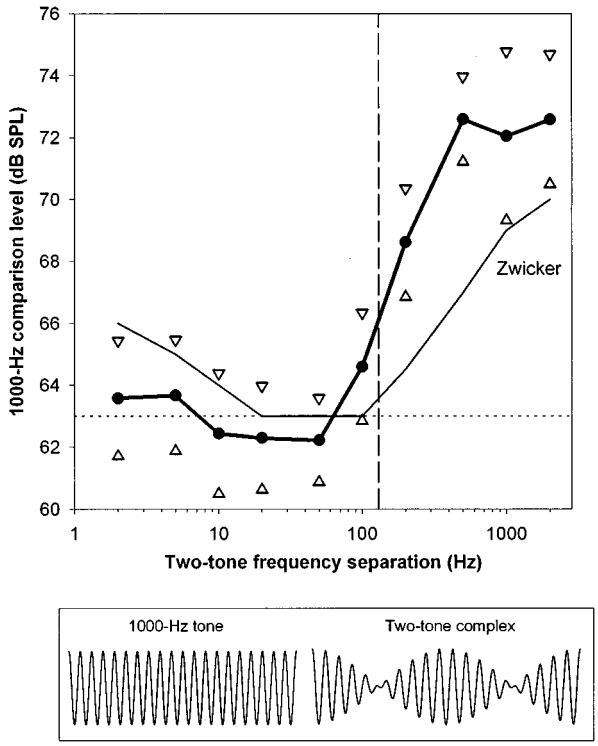


FIG. 1. Averaged loudness balance data in normal-hearing listeners between a 1000-Hz comparison tone (y axis) and a two-tone complex standard as a function of the frequency separation (x axis). The data were obtained from an adaptive procedure which tracked the 29% (regular triangles) and 71% (inverted triangles) louder response to the 1000-Hz tone on the psychometric function. The points of equal loudness (filled circles) were the average of the 29% and 71% response levels. Zwicker and Fastl's (1990) data are represented by the solid thin line. Examples of the waveforms are shown on the bottom panel.

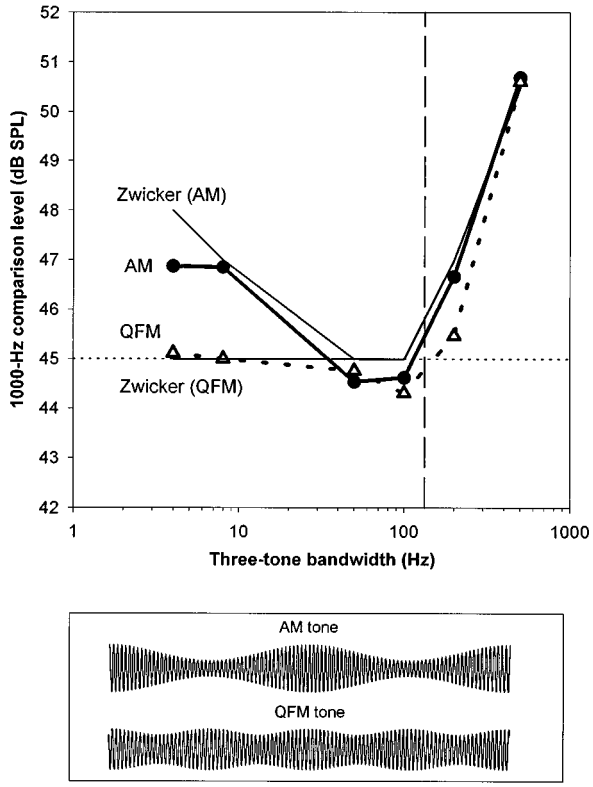


FIG. 2. Averaged loudness balance data in normal-hearing listeners between a 1000-Hz comparison tone (y axis) and a three-tone complex standard as a function of the total bandwidth (x axis). Filled circles represent data for the amplitude modulated (AM) stimuli and open triangles represent data for the quasi-frequency-modulated (QFM) stimuli. Zwicker and Fastl's (1990) data are represented by two solid thin lines. Examples of the waveforms are shown on the bottom panel.

in the AM stimulus than the QFM stimulus. The total bandwidth of the three-tone complex ($2\,f_m$) was selected at 4, 8, 50, 100, 200, and 500 Hz. The comparison stimulus was again the 1000-Hz tone.

B. Results and discussion

Since all four subjects had a similar pattern of results, only the averaged data for the two-tone loudness balance experiment are shown in Fig. 1. The filled circles represent the sound pressure level that was required for the 1000-Hz comparison (the y axis) to balance the loudness to the two-tone complex as a function of the frequency separation (the x axis). The upward and downward triangles are the averaged results from the lower and the upper sequences, which correspond to the 29% and 71% louder response to the 1000-Hz tone, and approximate the minus and plus one standard deviation from the mean. For comparison, the data from Zwicker and Fastl (1990, p. 189) are reproduced as the solid thin line. The horizontal dotted line represents the rms level (63 dB SPL) of the two-tone complex. The vertical broken line represents the critical band value of 133 at 1000 Hz (Glasberg and Moore, 1990).

Qualitatively, Zwicker and Fastl's results of the two-tone experiment are replicated by the present study using an adaptive procedure. When the frequency separation between the two tones is within the critical band, the rms intensity is added [$10\log(I+I)=3\text{ dB}+10\log I$] to produce a 3-dB effect on the loudness sensation of the two-tone complex (the dotted line). When the frequency separation is markedly wider than the critical band, the loudness is added [$I^{0.3}+I^{0.3}=(10*I)^{0.3}$] to produce a 10-dB effect ($10\log 10=10$ dB) on the loudness sensation of the two-tone complex. However, three deviations from Zwicker and Fastl's results are also apparent. First, for frequency separations below 10 Hz, Zwicker and Fastl observed a 6-dB effect which they attributed to the fact that loudness was determined by the peak level of the two-tone complex. The present data show only a 3.5-dB effect. Second, for frequency separations between 10 and 100 Hz, Zwicker and Fastl observed an intensity summation effect of precisely 3 dB. The present data show a 2-dB effect, suggesting that the two-tone complex actually sounds softer than the 1000-Hz tone at equal rms levels. Third, for frequency separations above the 133-Hz critical bandwidth, Zwicker and Fastl observed a shallow growth in loudness and a maximal 10-dB effect at the greatest frequency separation. In contrast, the present data show a much steeper loudness growth and a maximal 12-dB effect at the 1000-Hz separation.

The three-tone loudness balance data from the same four subjects are averaged and presented on the top panel of Fig. 2. The filled circles represent the loudness balance data for

the AM stimulus and the open triangles represent the data for the QFM stimulus. To avoid crowding the figure, the data from the upper and lower sequences are not plotted. The averaged difference between the 29% and 71% levels (approximating two standard deviations) across all bandwidth conditions is 3.0 dB for the AM stimulus and 2.4 dB for the QFM stimulus. Zwicker and Fastl's data (represented by two solid thin lines) and the present data are more consistent in this experiment than in the above two-tone experiment. Note the 2-dB difference between the AM and QFM stimuli for the 10-Hz or less total bandwidth conditions, which can be accounted for entirely by the 1.9-dB difference in the peak level between the AM and QFM stimuli. Note also the -0.5-dB difference between the present data and Zwicker and Fastl's data for three-tone bandwidths of 50 and 100 Hz. To summarize, although exp. 1 has replicated qualitatively Zwicker and Fastl's two- and three-tone experiments, severe deviations were also apparent, particularly in the two-tone experiment, where a much smaller peak-level effect, a less than perfect intensity summation effect, and a much steeper transition across the critical band were observed.

III. EXPERIMENT 2: LOUDNESS OF THE SAM NOISE IN ACOUSTIC HEARING

Experiment 1 demonstrated a temporal effect on loudness for slow temporal fluctuations. Because of the critical band analysis in normal hearing, a broadband stimulus is likely to be resolved into multiple bands, each of which will have less temporal fluctuations than the original overall fluctuations. For example, in the three-tone experiment, when the component frequency separation is much wider than the critical band, the ''brain'' perceives only the three individual components at the output of the periphery, with no difference in the temporal envelope between the AM and QFM stimuli. For this reason, and also because of the more dominant cross-critical-band loudness addition, the effect of the temporal envelope on loudness is difficult to observe in normal-hearing listeners except for the within-critical-band temporal fluctuations. In the present experiment, we used a sinusoidally amplitude-modulated (SAM) noise to circumvent the critical band analysis in the sense that the SAM noise effectively prevents spectral sidebands from being detected. Because of the continuous spectrum of the noise, amplitude modulation only increases the overall rms level, but does not change the spectral distribution of the noise, as in the case of the above three-tone experiment. We note, however, that the SAM noise does not prevent the reduction in modulation depth produced by the critical band analysis in normal-hearing listeners. A similar approach has been used to remove spectral cues in normal-hearing listeners in order to study exclusively the role of temporal cues in pitch perception (Burns and Viemeister, 1976) and in speech recognition (Schroeder, 1968; Van Tasell *et al.*, 1987; Rosen, 1992; Shannon *et al.*, 1995).

A. Stimuli

The comparison stimulus was white noise with a bandwidth of 20 Hz to 14 kHz. The standard stimulus was the SAM noise, which was constructed according to the formula: $S(t)=A[1+m\ \cos(2\pi f_m t)]n(t)$, where A was the carrier amplitude, m was the modulation index or depth, f_m was the modulation frequency, and $n(t)$ was the waveform of the white noise. Given fixed values of A and m, the rms level of the modulated noise was increased by a factor of $(1+m^2/2)^{0.5}$ relative to the unmodulated noise, whereas the peak amplitude was increased by a factor of $(1+m)$ (Viemeister, 1979). For example, in the case of a 100% amplitude modulation ($m=1.0$), the peak amplitude of the SAM noise was increased by 6 dB, whereas the rms level increased by only 1.8 dB. In other words, if the peak level determined loudness, a maximal difference of 4.2 dB in the rms level would be observed between the white noise and the SAM noise. The absolute thresholds for both the white noise and the SAM noise were at about 20 dB SPL. Three overall rms levels for the standard SAM noise were selected of 30, 45, and 70 dB SPL, corresponding to soft, comfortably loud, and loud sensations, respectively. Three modulation depths at 0.25, 0.5, and 1.0 were used for the 45-dB SPL condition, two depths of 0.5 and 1.0 for the 30 dB SPL, and only one depth of 1.0 for the 70 dB SPL. The modulation frequencies used were 4, 10, 20, 40, 60, 80, 100, 200, 400, and 1000 Hz.

B. Results and discussion

Figure 3 shows the averaged loudness balance data as a function of modulation frequency (x axis) and as a function of the overall level of the SAM noise (three panels). To have the same scale for the y axis in all three level conditions, the results are plotted as the difference in the rms level between the white noise and the SAM noise at the point of subjective equality in loudness. For example, if a 48-dB white noise is loudness balanced to a 45-dB SAM noise, then the y coordinate in Fig. 3 is 3 dB. The dotted line on each panel indicates no difference in the rms level between the SAM noise and the white noise at the point of subjective equality. The filled circles on all three panels represent the loudness balance data in the 100% modulation condition. The open squares represent the data obtained for 50% modulation (the upper and middle panels) and the open triangles represent the data for 25% modulation (the middle panel only). The measurement variability (the difference between the mean and the upper or the lower sequence) was consistent in all conditions and had an average value of 1.2 dB.

Three major points can be noted from Fig. 3. First, the temporal effect on the loudness is nonmonotonic as a function of the overall level. The greatest temporal effect on each panel changes from 2 dB at 30 dB SPL to 3.5 dB at 45 dB SPL to less than 1 dB at 70 dB SPL. The 3.5-dB magnitude of the temporal effect approaches the 4.2-dB maximum possible effect under the 100% modulation condition, if the peak level is presumed to determine the loudness. Second, the louder sensation produced by the amplitude-modulated noise, when present (30 and 45 dB SPL), occurs well beyond the 10-Hz range observed in Zwicker and Fastl's original two- and three-tone experiments. This louder sensation effect is relatively constant for modulation frequencies below 100, gradually rolls off from 100 to 1000, and eventually disappears at the 1000-Hz modulation frequency. This low-pass

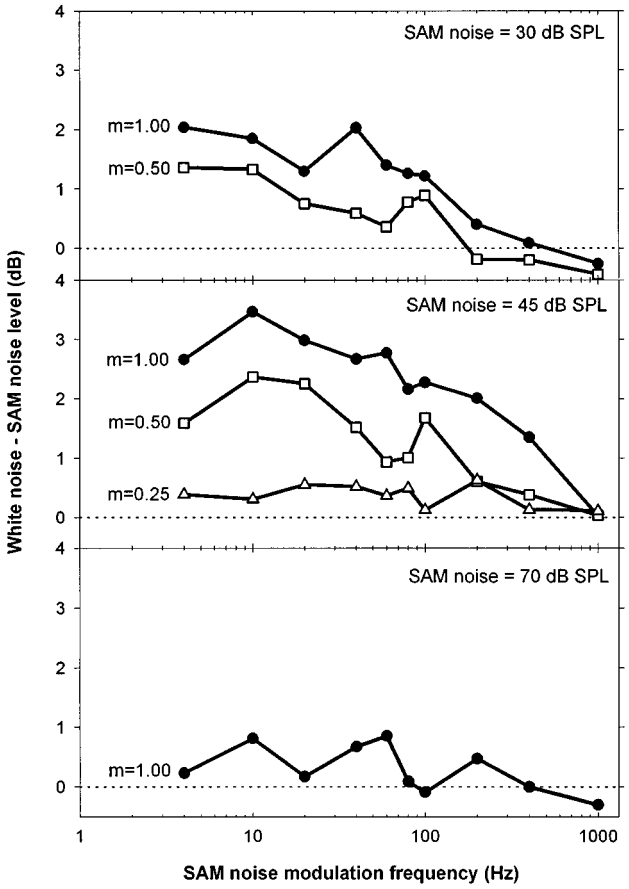


FIG. 3. Averaged loudness balance data in normal-hearing listeners represented by the rms level difference (y axis) between white noise and SAM noise as a function of modulation frequency (x axis), overall stimulus level (three panels), and modulation depth (100% = circles, 50% = squares, and 25% = triangles).

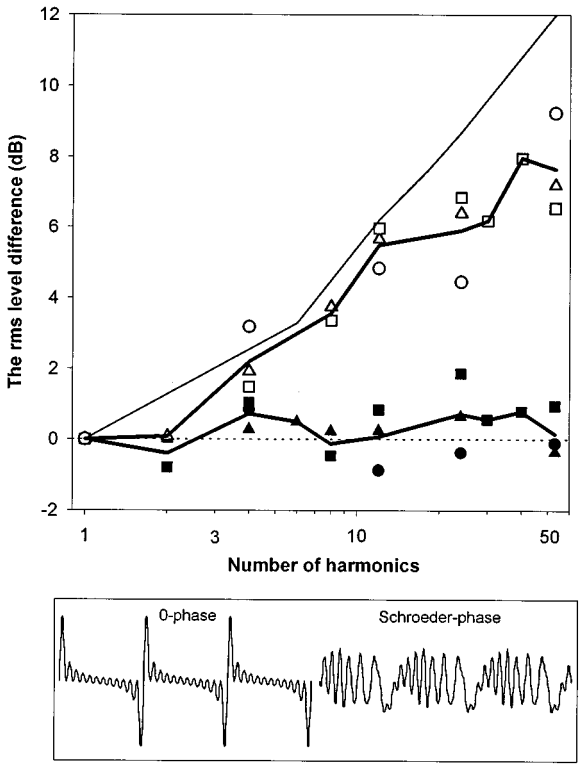


FIG. 4. Individual and averaged threshold and uncomfortable loudness data for three cochlear implant listeners. The data are represented by the rms level difference (y axis) between the zero- and Schroeder-phase stimuli as a function of number of harmonics (x axis). The individual data are represented by filled symbols for the threshold difference and open symbols for the uncomfortable level difference. The two thick lines represent the averaged data. The solid thin line represents the rms level difference between the zero- and Schroeder-phase stimuli when their peak level was equated. The dotted thin line indicates equal rms level for the two stimuli. Examples of waveforms (number of harmonics = 12) are shown in the bottom panel.

characteristic pattern of the temporal effect on loudness is qualitatively reminiscent of the temporal modulation transfer function measured by the modulation detection method (Viemeister, 1979). Third, the temporal effect on loudness was reduced significantly as the modulation depth decreased from 100% to 50% and to 25%. At 30 dB SPL for modulation frequencies between 4 and 200 Hz (the top panel), the 50% modulation produced an averaged temporal effect that was 0.7 dB lower than the 100% modulation ($t[df=7]=5.6$, $p<0.01$). At 45 dB SPL (the middle panel), the difference was 1.0 dB ($t[df=7]=3.7$, $p<0.01$) between the 100% and 50% modulations and was 0.43 dB ($t[df=7]=3.9$, $p<0.01$) between the 50% and 25% modulations.

IV. EXPERIMENT 3: LOUDNESS OF HARMONIC STIMULI IN ELECTRIC HEARING

To separate entirely the critical band analysis from the temporal envelope effects on loudness sensation of dynamic stimuli, we measured the loudness of two harmonic stimuli with the same magnitude spectrum but different phase spectra in three cochlear implant listeners. The different phase spectra used dramatically changed the temporal envelope or the peak level, but produced no change in the rms level. Direct stimulation of the auditory nerve bypassed the normal cochlear filtering processing, so that the overall temporal envelope of the harmonic stimuli was preserved in cochlear implant listeners.

A. Stimuli

Two types of phase manipulations were used in this experiment. One was called the zero-phase stimulus, in which all harmonics had the same 0° starting phase and the resulting waveform had the largest peak factor. The other was called the Schroeder-phase stimulus, in which the starting phase was manipulated to result in a waveform with the smallest peak factor (Schroeder, 1970). The Schroeder phase was defined as: $\theta_n=\pi n(n-1)/N$, where n is the harmonic number and N is the total number of harmonics. In this experiment, the fundamental frequency was 100 Hz and the number of harmonics ranged from 1 to 53. The bottom panel of Fig. 4 shows typical waveforms of the zero-phase and Schroeder-phase stimuli with 12 harmonics. Although the rms level was the same, the zero-phase stimulus had greater peak level than the Schroeder-phase stimulus.

TABLE I. Average threshold and uncomfortable loudness levels (ULL) in three cochlear implant listeners. The data were collected for both the zero-phase and Schroeder-phase (S-phase) stimuli as a function of the harmonic number, and are expressed as the rms level in μA.

Number of harmonics	Threshold		ULL	
	O-phase	S-phase	O-phase	S-phase
1	0.8	0.8	52.8	52.8
2	1.4	1.4	59.7	60.2
4	1.5	1.7	56.7	75.0
8	3.1	2.9	59.9	89.4
12	2.5	2.6	55.1	104.8
24	3.4	4.0	56.4	111.8
30	3.1	3.3	59.6	121.5
40	3.6	3.9	55.4	138.5
53	6.1	6.0	51.1	124.0

B. Procedure

In addition to the adaptive procedure in the loudness balance task, the method of limits was used to measure the threshold and uncomfortable loudness levels of the zero-phase and Schroeder-phase stimuli. In each trial of the threshold estimate, the level of the harmonic stimulus was adjusted by the listener to reach a criterion of just audible from a subthreshold level and was then adjusted to a criterion of just inaudible from a suprathreshold level. The just ''audible'' and ''inaudible'' levels were averaged to result in one estimate of the threshold. The uncomfortable level was estimated from only the ascending sequence in which the stimulus level was increased from soft to loud to uncomfortably loud. In a typical run, 10 to 12 such estimates were obtained. After determining the threshold and uncomfortable level for all harmonic conditions, the loudness balance task was performed from the threshold to the uncomfortable level between the zero- and Schroeder-phase stimuli for the 12-harmonic condition only.

C. Results and discussion

Table I shows both the zero-phase and the Schroeder-phase data for the threshold and uncomfortable loudness measures as a function of the number of harmonics. Notice the similar rms level between the zero- and Schroeder-phase stimuli for the threshold measure and the increasing difference between the two stimuli for the uncomfortable loudness measure. Another interesting point is that all zero-phase uncomfortable loudness levels have similar rms values. At present, we do not know the physiological mechanisms underlying these phenomena.

Figure 4 shows two theoretical predictions (Zeng and Shannon, 1995) and the difference in the threshold and uncomfortable levels between the zero- and Schroeder-phase stimuli. The solid thin line represents the ''equal-peak, equal-loudness'' prediction corresponding to the calculated rms level difference between the zero-phase and Schroeder-phase stimuli when their peak amplitudes are made equal. This rms level difference under the ''equal-peak'' condition increases monotonically from 0 to 6 and 12 dB as the number of harmonics is increased from 1 to 12 and 53, respectively. On the other hand, the dotted thin line represents the

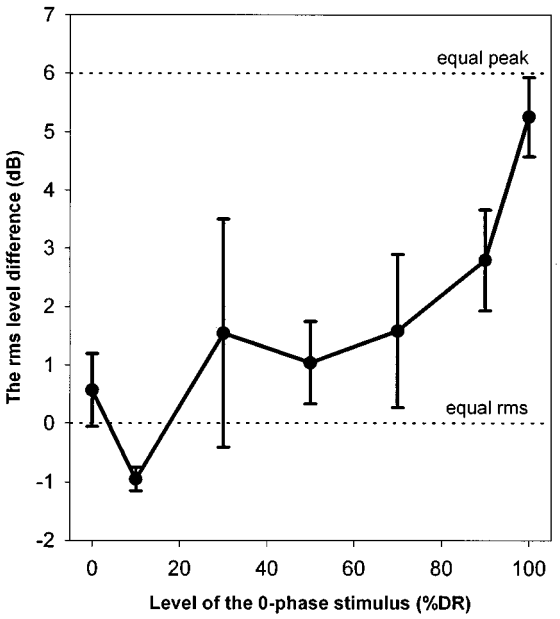


FIG. 5. Averaged loudness balance data between the zero-phase and Schroeder-phase stimuli for the 12-harmonic condition in three cochlear implant listeners. The balance data are represented by the rms level difference (y axis) between the zero- and Schroeder-phase stimuli as a function of the standard zero-phase stimulus level plotted as a percentage of dynamic range (x axis). The ''equal rms'' dotted thin line shows predictions of the ''equal-rms, equal-loudness'' hypothesis, while the ''equal peak'' dotted thin line shows predictions of the ''equal-peak, equal-loudness'' hypothesis.

''equal-rms, equal-loudness'' prediction, indicating no temporal envelope effect on loudness. The filled symbols represent the individual listener's threshold difference in the rms level between the zero-phase and Schroeder-phase stimuli and the solid thick line through these filled symbols represents the averaged results. Correspondingly, the open symbols represent the individual data for the uncomfortable level and the solid thick line through these open symbols represents the averaged results. It is obvious from Fig. 4 that the threshold data are consistent with the ''equal-rms, equal-loudness'' hypothesis, whereas the uncomfortable loudness data are significantly different from this ''equal-rms'' hypothesis. The trend of the uncomfortable loudness data is more closely predicted by the ''equal-peak, equal-loudness'' hypothesis, despite a greater deviation from the prediction as the number of harmonics was increased. In general, the present data are consistent with and extend the findings of Zeng and Shannon (1995). Both sets of data suggest that the temporal effect of loudness in electric hearing is level dependent, wherein the threshold is determined by the rms level and the uncomfortable loudness is determined more by the peak level of the temporal envelope.

To study the transition of this level dependence from the threshold to the uncomfortable loudness level, loudness balance data were obtained between the zero-phase and Schroeder-phase stimuli for the 12-harmonic complex only. Figure 5 shows the rms difference between the zero- and Schroeder-phase stimuli (y axis) as a function of the standard zero-phase level, which has been normalized according to each individual listener's dynamic range (x axis). The lower

dotted line represents the ''equal-rms, equal-loudness'' prediction, while the upper dotted line represents the ''equal-peak, equal-loudness'' prediction, corresponding to a maximal 6-dB difference in the rms level between the zero-phase and Schroeder-phase stimuli (see Fig. 4). Figure 5 shows that as the standard level increases from threshold (0% dynamic range) to uncomfortable loudness (100% dynamic range), the rms-level difference between the two stimuli increases monotonically from 0 to 5.3 dB. We have not attempted to model this loudness balance function, but this monotonic level dependence of the temporal effect on loudness in electric hearing is in great contrast to the nonmonotonic effect in acoustic hearing.

V. FINAL REMARKS

Traditional views on the temporal effect on loudness have been limited to slow changes in the dynamic stimuli; for example, the beats in a two-tone complex (Zwicker and Fastl, 1990). Recently, attention has been paid to the dynamics of loudness sensation as a result of much faster changes in two-tone noise complexes (Hellman, 1985), in the short-term temporal energy distribution (Stecker and Hafter, 1996), and in the temporal envelope of harmonic complexes (Carlyon and Datta, 1997) and amplitude-modulated tones (Moore *et al.*, 1997). The present work studied the loudness of dynamic stimuli by measuring the loudness of a sinusoidally amplitude-modulated (SAM) noise in normal-hearing listeners and the loudness of zero-phase versus Schroeder-phase stimuli in cochlear implant listeners. The results clearly demonstrate that fast temporal fluctuations up to 400 Hz can increase the loudness of dynamic stimuli. The increase in loudness has a low-pass characteristic qualitatively similar to that of the temporal modulation transfer function (Viemeister, 1979). This low-pass pattern suggests that the temporal effect on loudness is not determined by the temporal integration, as suggested by Zwicker and Fastl (1990), which has a time constant of about 100 ms (e.g., Zwislocki, 1960), but rather is determined by the absolute temporal resolution, which has a time constant 1–2 orders of magnitude lower (Viemeister, 1996).

The results of the present two- and three-tone experiments are generally, although not entirely, consistent with previous results (Zwicker and Fastl, 1990; Moore *et al.*, 1997). At low-modulation frequencies, Zwicker and Fastl observed a maximum 3-dB ''peak-level'' effect for both the two- and three-tone experiments. The present study observed only a 0.5-dB effect for the two-tone experiment and a 2-dB effect for the three-tone experiment. Moore *et al.* observed an essentially 0-dB effect for the 4-Hz modulation frequency (i.e., the rms level determines loudness). For modulation frequencies between 10 Hz and the critical bandwidth, Zwicker and Fastl found that the loudness was equal when the rms level was equal between the tone and the modulated stimuli; on the contrary, the present and Moore *et al.* studies found that the modulated stimuli were actually softer than the pure tone (about 1-dB effect) at equal rms levels. Since the present study used an adaptive procedure, Moore *et al.* used a ''bracket'' loudness balance procedure, and Zwicker and Fastl used an unreported procedure, the discrepancy among the three studies might be due to the procedural difference and/or other unknown factors.

Another important difference is noted between the two- or three-tone experiments and the SAM noise experiment: the experiments with sinusoidal carriers generally produced a relatively small and somewhat inconsistent temporal effect on loudness sensation, whereas the present experiment with a noise carrier demonstrated a consistently louder sensation for the amplitude-modulated noise than the unmodulated noise at equal rms levels, at least for low overall stimulus levels and over a large modulation frequency range. This difference in the temporal effect on loudness appears to be related to the nature of narrow-band versus wideband stimuli. Carlyon and Datta (1997) found that for five harmonics centered at 1100 Hz there was little difference in loudness between the positive and negative Schroeder-phase stimuli at equal rms levels (see also Preece and Wilson, 1988), but a significant phase effect occurred when the same five harmonics were presented simultaneously with a large number of ''off-frequency'' components. Carlyon and Datta attributed their findings to a fast-acting compression and interactions among these components at the basilar membrane level, which attenuated the ''peakier'' auditory filter output of the positive-phase stimulus more than the ''smoother'' output of the negative-phase stimulus (Kohlrausch and Sander, 1995). Both the present and Carlyon and Datta's studies suggest that wideband stimuli may be needed to illustrate reliably the temporal effect on loudness sensation.

The present study also demonstrates a level dependency of the temporal effect on the loudness of dynamic stimuli in both acoustic and electric hearing, but in opposite directions. In normal-hearing listeners, the modulated noise was louder than the unmodulated noise at low rms levels (30 and 45 dB SPL) but produced essentially equal loudness at a higher rms level of 70 dB SPL. This result suggests that loudness is determined more by the peak level at low sensation levels and more by the rms level at high sensation levels. Conversely, the result obtained from cochlear implant listeners stimulated with zero- and Schroeder-phase stimuli suggests that loudness is determined more by the rms level at low sensation levels and more by the peak level at high sensation levels. This apparent discrepancy between normal-hearing and implant listeners may be explained partially by the fast-acting basilar membrane mechanics which behaves more linearly at low levels and becomes compressive at medium and high levels (Ruggero, 1992; Oxenham and Plack, 1997), and partially by the expanding exponential loudness growth with direct stimulation of the auditory nerve (Zeng and Shannon, 1992, 1994). In normal-hearing listeners, the peak–trough difference in the modulated stimuli would be preserved in the low-level linear region and greatly reduced in the high-level compressive region, which can at least qualitatively produce the observed greater loudness effect at low levels. On the other hand, if we assume that the exponential loudness growth in electric stimulation is due to a ''fast-acting'' expansion, then the observed reversed level-dependent loudness effect in cochlear implant listeners could be explained. Thus the compression of the cochlea could explain the nor-

mal pattern of results, while the loss of that cochlear compression, coupled with the central expansion, could reverse the pattern in implants. Quantitatively, Moore *et al.* (1997) calculated the ''post-compression'' rms and peak-level difference between the modulated and unmodulated stimuli and found that a measure between the rms and peak level (after compression) could explain their results. We wonder whether a phenomenological model incorporating a peripheral compression and a central expansion (Zeng and Shannon, 1994; Zeng *et al.*, 1997) would produce such a measure that can better predict the present data and the Moore *et al.* (1997) data.

Although the physiological processes underlying the loudness sensation of dynamic stimuli remain unclear, the present results have direct clinical relevance to the design of the compression circuit in hearing aids and cochlear implants. If the purpose of the compression circuit is to restore normal loudness in hearing-impaired listeners, then the design should take into account the temporal fluctuation or the peak amplitude of dynamic stimuli such as speech sounds. For example, to restore normal loudness of speech sounds in cochlear implant listeners, the compression circuit should prescribe the gain for the soft consonants according to their short-term rms level while compressing the loud vowels according to their peak level. In other words, this compression circuit will need two different time constants that are dependent on the input sound level. At or near the threshold level, a long-term time constant (75–100 ms) is needed to take the temporal integration into account, whereas for loud sensations, a short or zero time constant is needed to compress the peak level of the speech sound. Investigations are under way to evaluate such a design strategy to see whether either improved speech recognition, better speech quality, or both can be achieved for hearing-impaired listeners.

ACKNOWLEDGMENTS

The authors would like to acknowledge John J. Galvin III, Lendra Friesen, and Qian-Jie Fu for their help in data collection and in programming experiments. The authors also thank Bert Schlauch, Bob Shannon, Alena Wilson, Chris Plack, and Brian Moore for their comments on this paper. This work was supported by the National Institutes of Health (NIDCD-DC02267).

Boothroyd, A., Erickson, F. N., and Medwetsky, L. (**1994**). ''The hearing aid input: a phonemic approach to assessing the spectral distribution of speech,'' Ear Hear. **15**, 432–442.

Burns, M. E., and Viemeister, N. F. (**1976**). ''Nonspectral pitch,'' J. Acoust. Soc. Am. **60**, 863–869.

Byrne, D., and Dillon, H. (**1986**). ''The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid,'' Ear Hear. **7**, 257–265.

Carlyon, R. P., and Datta, A. J. (**1997**). ''Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system,'' J. Acoust. Soc. Am. **101**, 3636–3647.

Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65.

Fletcher, H. (**1953**). *Speech and Hearing in Communication* (Van Nostrand, New York), p. 87.

Fletcher, H., and Munson, W. A. (**1933**). ''Loudness, its definition, measurement, and calculation,'' J. Acoust. Soc. Am. **5**, 82–108.

Fletcher, H., and Munson, W. A. (**1937**). ''Relation between loudness and masking,'' J. Acoust. Soc. Am. **9**, 1–10.

Florentine, M., and Zwicker, E. (**1979**). ''A model of loudness summation applied to noise-induced hearing loss,'' Hearing Res. **1**, 121–132.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hearing Res. **47**, 103–138.

Green, D. M. (**1964**). ''Psychoacoustics and detection theory,'' in *Signal Detection and Recognition by Human Observers*, edited by J. A. Swets (Wiley, New York), pp. 58–94.

Hellman, R. P. (**1985**). ''Perceived magnitude of two-tone-noise complexes: Loudness, annoyance, and noisiness,'' J. Acoust. Soc. Am. **77**, 1497–1504.

ISO-Rec. 532 (**1975**). ''Acoustics-method for calculating loudness level'' (International Organization for Standardization, Geneva).

Jesteadt, W. (**1980**). ''An adaptive procedure for subjective judgments,'' Percept. Psychophys. **28**, 85–88.

Kohlrausch, A., and Sanders, A. (**1995**). ''Phase effects in masking related to dispersion in the inner ear II. Masking period patterns of short targets,'' J. Acoust. Soc. Am. **97**, 1817–1829.

Launer, S. (**1995**). '' Loudness perception in listeners with sensorineural hearing impairment,'' Ph.D. thesis, Olderburg, Germany.

Launer, S., and Bachler, H. (**1996**). ''Implications of modeling sensorineural loss to hearing instrument design, selection, and fitting,'' abstract from ''Issues in advanced hearing aid research'' p. 18, Lake Arrowhead, California.

Moore, B. C. J., and Glasberg, B. R. (**1996**). ''A revision of Zwicker's loudness model,'' Acustica **82**, 335–345.

Moore, B. C. J., and Glasberg, B. R. (**1997**). ''A model of loudness perception applied to cochlear hearing loss,'' Aud. Neurosci. **3**, 289–311.

Moore, B. C. J., Wojtczak, M., and Vickers, D. A. (**1996**). ''Effect of loudness recruitment on the perception of amplitude modulation,'' J. Acoust. Soc. Am. **100**, 481–489.

Moore, B. C. J., Launer, S., Vickers, D. A., and Baer, T. (**1997**). ''Loudness of modulated sounds as a function of modulation rate, modulation depth, modulation waveform, and overall level,'' Proceedings of the 11th International Symposium on Hearing, Grantham, England (in press).

Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.

Preece, J. P., and Wilson, R. H. (**1988**). ''Detection, loudness, and discrimination of five-component tonal complexes differing on crest factor,'' J. Acoust. Soc. Am. **84**, 166–171.

Rosen, S. (**1992**). ''Temporal information in speech: acoustic, auditory and linguistic aspects,'' Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.

Ruggero, M. A. (**1992**). ''Responses to sound of the basilar membrane of the mammalian cochlea,'' Current Opinion. Neurobiol. **2**, 449–456.

Schlauch, R. S., and Wier, C. C. (**1987**). ''A method for relating loudness matching and intensity discrimination data,'' J. Speech Hear. Res. **30**, 13–20.

Schroeder, M. R. (**1968**). ''Reference signal for signal quality studies,'' J. Acoust. Soc. Am. **44**, 1735–1736.

Schroeder, M. R. (**1970**). ''Synthesis of low-peak-factor signals and binary sequences with low autocorrelation,'' IEEE Trans. Inf. Theory **16**, 85–89.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.

Stecker, G. C., and Hafter, E. R. (**1996**). ''An effect of temporal energy distribution on loudness,'' J. Acoust. Soc. Am. **100**, 2627.

Stevens, S. S. (**1957**). ''On the psychophysical law,'' Psychol. Rev. **64**, 153–181.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., and White, R. L. (**1987**). ''Speech waveform envelope cues for consonant recognition,'' J. Acoust. Soc. Am. **82**, 1152–1161.

Viemeister, N. F. (**1979**). ''Temporal modulation transfer functions based upon modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380.

Viemeister, N. F. (**1996**). ''Auditory temporal integration: What is being accumulated?,'' Current Direct. Psychol. Sci. **5**, 28–32.

Vurek, L. S. *et al.* (**1981**). ''Opto-isolated stimulators used for electrically evoked BSER,'' Ann. Otol. Rhinol. Laryngol. **90**, Suppl. 82, 21–24.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). ''Better speech reception with cochlear implants,'' Nature (London) **352**, 236–238.

Zeng, F.-G., and Shannon, R. V. (**1992**). ''Loudness balance between electric and acoustic stimulation,'' Hearing Res. **60**, 231–235.

Zeng, F.-G., and Shannon, R. V. (**1994**). ''Loudness-coding mechanisms

inferred from electric stimulation of the human auditory system,'' Science **264**, 564–566.
Zeng, F.-G., and Shannon, R. V. (**1995**). ''Loudness of simple and complex stimuli in electric hearing,'' Ann. Otol. Rhinol. Laryngol. **104**, 235–238.
Zeng, F.-G., and Turner, C. W. (**1991**). ''Binaural loudness matches in unilaterally impaired listeners,'' Q. J. Exp. Psychol. **43**A, 565–583.
Zeng, F.-G., Shannon, R. V., and Hellman, W. S. (**1997**). ''Physiological processes underlying psychophysical laws,'' Proceedings of the 11th International Symposium on Hearing, Grantham, England (in press).
Zwislocki, J. (**1960**). ''Theory of temporal auditory summation,'' J. Acoust. Soc. Am. **32**, 1046–1060.
Zwicker, E., and Fastl, H. (**1990**). *Psychoacoustics—Facts and Models* (Springer-Verlag, Berlin).
Zwicker, E., and Scharf, B. (**1965**). ''A model of loudness summation,'' Psychol. Rev. **72**, 3–26.
Zwicker, E., Flottorp, G., and Stevens, S. S. (**1957**). ''Critical band width in loudness summation,'' J. Acoust. Soc. Am. **29**, 548–557.

The perception of frequency peaks and troughs in wide frequency modulations. IV. Effects of modulation waveform

Laurent Demany and Sylvain Clément
Laboratoire d'Audiologie Expérimentale et Clinique, Université Bordeaux 2, BP 63, 146 rue Leo-Saignat, F-33076 Bordeaux Cedex, France

(Received 11 February 1997; revised 21 July 1997; accepted 21 July 1997)

This work extends previous studies on the perceptual asymmetry between the local maxima and minima of wide frequency modulations (FMs) [L. Demany and K. I. McAnally, J. Acoust. Soc. Am. **96**, 706–715 (1994); L. Demany and S. Clément, J. Acoust. Soc. Am. **97**, 2454–2459 (1995); L. Demany and S. Clément, J. Acoust. Soc. Am. **98**, 2515–2523 (1995)]. In experiment 1, subjects had to discriminate frequency shifts in the temporally central vertex of V- and Λ-shaped FMs imposed on 200-ms sinusoidal tone bursts. The precise shapes of these FMs varied in eight steps from quasi-triangles (with a durationless central vertex) to quasi-squares (with a long-duration central vertex). The central vertex was either a minimum or a maximum, but in each case the corresponding frequency was near 1000 Hz and the FM span was about 0.5 oct. For each FM shape, the discrimination threshold was lower when the vertex was a maximum than when it was a minimum, but (in four subjects out of five) this difference decreased monotonically as the FM became less and less triangular. FM shape had a remarkably small effect on the discrimination of the maxima, and the thresholds measured for the sharpest maxima were unexpectedly low. In subsequent experiments, subjects had to discriminate frequency shifts in the starting point or the final point of unidirectional FMs (tone glides) that spanned about 0.5 oct in 100 ms. The relevant frequency extremum was near 1000 Hz in each condition. At the final point of the glides, discrimination was better for rising glides than for falling glides. At the starting point of the glides, discrimination was better for falling glides than for rising glides. Thus discrimination was always better when the relevant frequency extremum was a maximum than when it was a minimum, and this effect was produced both ''forward'' and ''backward.'' The latter fact suggests that the perceptual asymmetry of FM originates at least partly from central factors. © *1997 Acoustical Society of America.* [S0001-4966(97)02211-X]

PACS numbers: 43.66.Fe, 43.66.Hg, 43.66.Mk [WJ]

INTRODUCTION

This is the fourth in a series of papers devoted to the perception of ''instantaneous pitch'' in widely frequency modulated sounds (Demany and McAnally, 1994; Demany and Clément, 1995a, 1995b; henceforth, these three previous articles will be referred to as papers 1, 2, and 3). The whole series is focused on an intriguing perceptual asymmetry that may be termed ''the peak/trough effect.'' Our first experiment displaying this effect used periodic modulators consisting of the sum of a few sinusoids (with a maximum frequency of 10.5 Hz). These complex FMs were imposed on a sinusoidal carrier of about 1 kHz. Within such stimuli, most listeners appeared to hear a repetitive sequence of tones with precise pitches. As could be expected, the heard tones corresponded to local frequency extrema of the FM waveforms. However, although the FM waveforms were symmetric on the dimensions of time and log frequency, we found that tones were much more often heard at local maxima (FM peaks) than at local minima (FM troughs). This occurred even for maxima and minima at the same frequency.

In subsequent experiments, we simplified the FM waveforms. They were reduced to a single cycle of a cosine function on a logarithmic frequency scale. The subjects' task was then to discriminate stimuli that differed only by the frequency vertex occurring in their temporal centers (cf. Fig. 3 of paper 1). The cosine cycle of the FM started and ended either at 0° (in the ''trough'' condition) or at 180° (in the ''peak'' condition). We found that frequency shifts of a given vertex were typically much better detected in the ''peak'' condition than in the ''trough'' condition. This was the case for FM carriers consisting of sinusoids at low or medium frequencies, various kinds of harmonic complexes, and even amplitude-modulated noise. As assessed in this way, the peak/trough effect appeared to be a robust phenomenon: It did not diminish with practice in the discrimination task.

The goal of the present experiments was to investigate more systematically the effect of modulation waveform on the perception of FM peaks and troughs. We thus measured again, for new FM waveforms, just-noticeable shifts in frequency extrema. The FMs used in experiment 1, like the cosine modulators that we had previously employed, produced a single ''there and back'' pitch motion. However, instead of being cosinusoidal, their waveforms ranged from quasi-triangles (without *steady* vertex) to quasi-rectangles (with a long-duration steady vertex). In experiments 2 and 3, the FMs of primary interest were unidirectional rather than bidirectional. The frequency extrema under study were then located at the starting point or the end point of stimuli con-

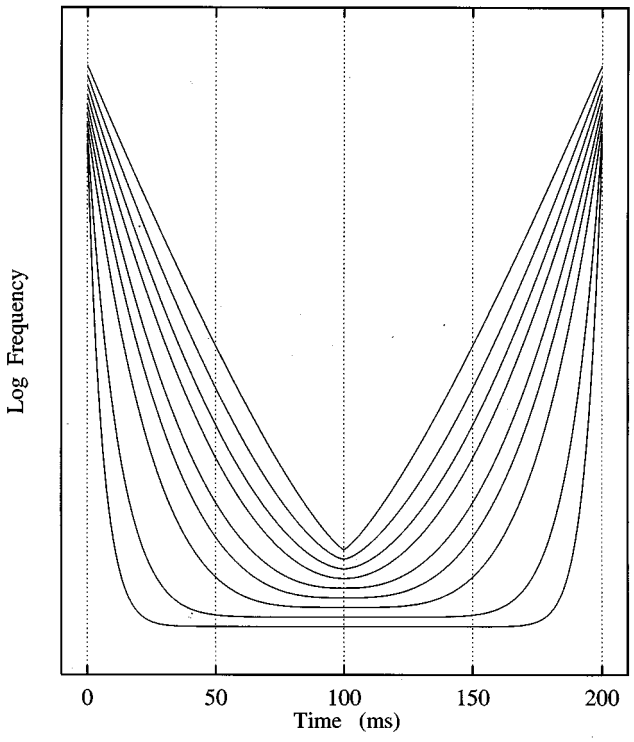


FIG. 1. The family of FM waveforms used in the ''trough'' condition of experiment 1. The vertical separations of the waveforms are arbitrary. In the standard stimuli, the initial and final instantaneous frequency was always 0.5 oct away from the temporally central frequency vertex.

sisting of a constantly rising or constantly falling frequency glide.

I. EXPERIMENT 1

A. Stimuli and rationale

In each trial of this experiment, the subject had to discriminate between two slightly different FMs imposed on sinusoidal tone bursts lasting 200 ms (complete envelope) and gated on and off with 10-ms cosinusoidal amplitude ramps. Each of the two FMs produced a temporally symmetric ''there and back'' pitch motion. For the ''standard'' stimulus, the frequency vertex reached after 100 ms (F_{vertex}) was at 1000 Hz and the initial and final value of instantaneous frequency (f_{inst}) was 0.5 oct away from F_{vertex}. The ''target'' stimulus, to be discriminated from the standard, differed from it by a slight shift of F_{vertex}, the initial and final value of f_{inst} being unchanged; this shift of F_{vertex} always increased the overall span of f_{inst}.

Figure 1 displays the family of standard FM waveforms used in the ''trough'' condition. (The vertical separations of the curves are merely intended to facilitate visual comparisons.) Upside down, this figure shows instead the standard FM waveforms of the ''peak'' condition. Within each stimulus, f_{inst} varied according to the formula:

$$f_{\text{inst}}(t) = F_{\text{vertex}} \exp(\alpha |t - 0.1|^{\eta}), \quad (1)$$

where t represents time in seconds ($0 \leqslant t \leqslant 0.2$). Across trial blocks, the exponent η took nine different values, ranging from 1.25 to 16. This parameter determined the sharpness of the frequency vertex occurring at $t = 0.1$. The vertex was extremely sharp for $\eta = 1.25$ (highest curve in Fig. 1). On the contrary, for $\eta = 16$ (lowest curve), f_{inst} was very close to F_{vertex} during more than half of the stimulus. Variable α (positive in the ''trough'' condition and negative in the ''peak'' condition) was adjusted as a function of η and F_{vertex} in order to obtain the required value of f_{inst} for $t = 0$ and $t = 0.2$, that is 707.1 Hz in the ''peak'' condition and 1414.2 Hz in the ''trough'' condition.

For each value of η, therefore, discrimination thresholds for F_{vertex} were measured in the ''peak'' and ''trough'' conditions. Precise predictions about the effect of η on thresholds could be drawn from a model presented by Horst (1989). We wished to compare our data to these predictions. For the largest value of η, given that f_{inst} was very close to F_{vertex} for a long time, it could be expected that thresholds would be lowest and that no clear peak/trough effect would be apparent: Obviously, it should become impossible to observe a peak/trough effect when η is so large that the standard stimuli of the ''peak'' and ''trough'' conditions differ from each other only during a few ms, at their very beginning and end. An interesting question was: As η decreases and gets closer and closer to 1, does the peak/trough effect always increase, or is there some intermediate value of η for which this effect is maximum? The second alternative, a nonmonotonic effect of η on the magnitude of the peak/trough effect, could seem more likely than the first one. To see why, note first that shifting F_{vertex} produced a concomitant shift in the FM slopes before and after $t = 0.1$. Shifts in an FM slope are much less detectable than frequency shifts in a steady tone (cf., e.g., Dooley and Moore, 1988). Thus, when η was large or very large, one could reasonably expect that subjects would perform the discrimination task by detecting shifts in the frequency plateau centered on $t = 0.1$; it was unreasonable to hypothesize instead that subjects would use an FM slope cue. When η was close to 1, however, the latter hypothesis became quite plausible because: (1) the FM waveforms did not contain any plateau centered on $t = 0.1$; (2) within each stimulus, the FM slopes were approximately constant for a relatively long time. Assume that FM slope becomes the *only* discrimination cue when η is close to 1. In this case, similar results are expected in the ''peak'' and ''trough'' conditions since the corresponding standard stimuli do not differ from each other with respect to FM slope (in oct/s). Even if one-half of the stimuli perceptually dominates the other half, the results should still be similar in the two conditions, because just-noticeable shifts in the slope of frequency glides appear to be quite similar for rising versus falling glides (Dooley and Moore, 1988).

B. Procedure

The stimuli were digitally generated in real time, at a sampling rate of 20 kHz, with the equipment already described in paper 1. Each FM waveform was defined in a table containing 1024 frequency samples (i.e., 5.12 frequency samples per ms). The output of the DAC was low-pass filtered at 8 kHz and presented binaurally, at 55 dB SPL, via

TDH 39 earphones. Five subjects with normal hearing, including the two authors, were tested individually in a double-walled soundproof booth.

On each trial, the standard and the target (presented in this order or the opposite order, at random) were separated by a 500-ms silent interval. The subject's identification of the target's position was given by pressing one of two keys, and was immediately followed by visual feedback. During each trial block, a threshold was measured for a fixed value of η. The adaptive procedure used to measure thresholds (for details, see paper 1, Sec. II A1) estimated the 75% correct point of the psychometric function. Within the experimental sessions, as well as preliminary training sessions, the ''peak'' and ''trough'' conditions were always presented in alternation. The training period was short (1–4 h, depending on the subject), but each subject had previously taken part in at least one closely related experiment.

Two ''sub-experiments'' were run. In the first one, η did not exceed 4; we did not use larger η values because we guessed that the peak/trough effect would already be quasi-inexistent for $\eta=4$. As the results showed that this was not the case, we performed (on only four of the five subjects) a second sub-experiment in which η was equal to 4, 8, or 16.

C. Results

The individual results of the five subjects are displayed in five panels of Fig. 2. In these panels, each symbol represents the mean of ten threshold measurements. The results obtained for $\eta=4$ in the two sub-experiments are shown separately. Note that the two axes of the figure are logarithmic.

Four of the five subjects (see the four upper panels of Fig. 2) behaved similarly. For these subjects, thresholds were essentially a monotonic function of η, in both the ''peak'' and ''trough'' conditions. However, the effect of η on thresholds was larger in the ''trough'' condition than in the ''peak'' condition. For the largest value of η, thresholds were lowest and there was only a weak peak/trough effect. The peak/trough effect increased monotonically as η decreased, and became quite pronounced for $\eta=1.25$. For subject JJ, by contrast, the thresholds measured in the first sub-experiment did not strongly depend on η and the peak/trough effect was always weak. This subject was not tested in the second sub-experiment.

D. Discussion

Overall, the results did not confirm our expectation that the peak/trough effect would be a nonmonotonic function of η. Since this expectation was based on the assumption that, in both the ''peak'' and ''trough'' conditions, the subjects' discrimination cue for the lowest value(s) of η would be FM slope rather than F_{vertex}, one is led to think that this assumption was wrong. Indeed, as subjects of the experiment, the authors felt that the discrimination cue that they used in the ''peak'' condition was the pitch corresponding to F_{vertex} for each value of η, even the lowest one. (It was more difficult to specify from introspection the discrimination cue, or cues, used in the ''trough'' condition.) Moreover, *all* the thresholds measured in the ''peak'' condition were too low to be consistent with the FM slope hypothesis: According to Dooley and Moore (1988), the Weber fraction for the detection of shifts in FM slope is about 0.06; this would yield a threshold of 36 cents in the present experiment. Note that although all the thresholds measured in the ''peak'' condition were much lower, this was not the case in the ''trough'' condition: In the latter condition, for $\eta=1.25$, the mean threshold of the five subjects was 33 cents. Thus FM slope may have been the dominant discrimination cue in the ''trough'' condition when η was very small.

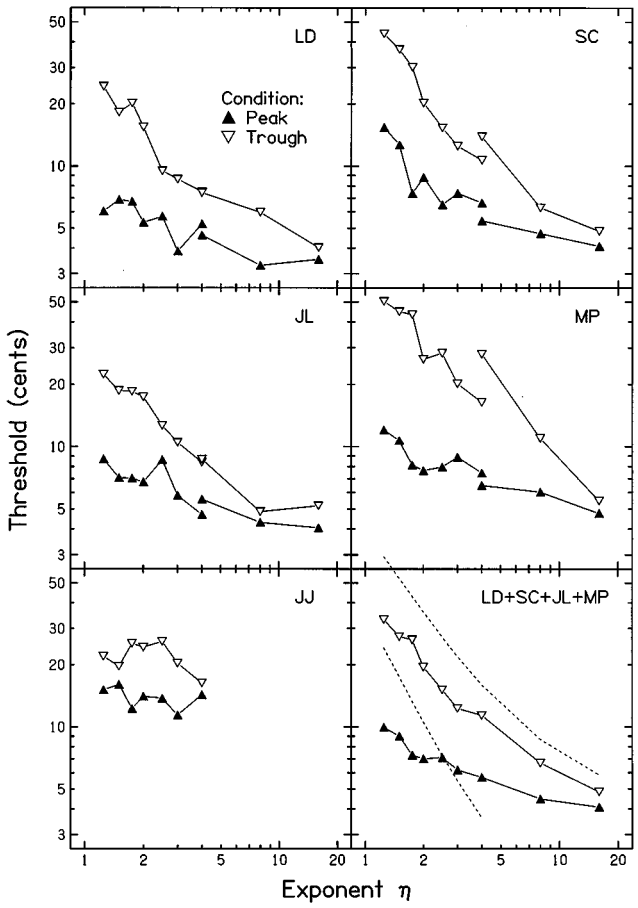


FIG. 2. Results of experiment 1. The discrimination thresholds of the five subjects (LD, SC, JL, MP, JJ) are displayed in separate panels, as a function of the nature of the frequency vertex (peak or trough) and the value of exponent η in Eq. (1). The axes are logarithmic. Thresholds are expressed in ''cents'' rather than in Hz for the sake of consistency with papers 1–3. One cent = 1/100 semitone = 1/1200 oct. For a reference frequency of 1000 Hz (the standard value of F_{vertex}), 10 cents correspond to a frequency difference of about 5.8 Hz. In the bottom-right panel (mean results for subjects LD, SC, JL, and MP), the two dashed curves represent predictions of a model described by Horst (1989; see the text for details).

The bottom-right panel of Fig. 2 displays the average results of the four subjects who behaved similarly. The two dashed curves are predictions from the model proposed by Horst (1989) to account for the results of his own study on the discrimination of FM waveforms. In Horst's study, the standard and target stimuli also consisted of a ''there and back'' frequency movement, but the central frequency vertex was always a peak. Horst assumed that subjects detect shifts in F_{vertex} by making measurements of its value within a rect-

TABLE I. Width (2τ) of the temporal window minimizing S in Eq. (2), as a function of η; $a=16$.

η	1.25	1.5	1.75	2.0	2.5	3.0	4.0	8.0	16.0
2τ (ms)	8.0	10.2	12.8	15.4	20.8	26.4	37.2	72.8	112.4

angular temporal window centered on the vertex. The accuracy of these frequency measurements is limited by two factors: (1) the variation of f_{inst} within the window; (2) a reciprocal relation between frequency accuracy and time, reflecting the well-known ''uncertainty principle'' (see Hartmann, 1997, Chap. 13). The first factor favors the choice of a narrow window, but the second factor favors instead wide windows. Subjects are supposed to adopt the best compromise, that is to select the window width (2τ) that minimizes the sum of the imprecisions due to the two factors. This sum (S) can then be considered as an approximation of the detection threshold for shifts in F_{vertex}. For our stimuli, following Horst's model, S was given (in Hz) by the following equation:

$$S = F_{\text{vertex}}[\exp(\alpha \cdot \tau^{\eta}) - 1] + (2\tau\ a)^{-1}, \tag{2}$$

where $F_{\text{vertex}}=1000$ and a is a constant which can be assessed from psychophysical data concerning the effect of duration on the frequency discrimination of tone bursts near 1000 Hz. In his own equation for S, Horst set a at 2.5. Taking the same value of a in Eq. (2) leads to the predictions corresponding to the upper dashed curve in the bottom-right panel of Fig. 2. However, the psychophysical data reported by Moore (1973) suggest that it is much more realistic to set a at 16: For a 1000-Hz tone burst of 25 ms, Moore measured a frequency discrimination threshold of about 2.5 Hz, and this threshold was multiplied by about 2 and 4, respectively, when stimulus duration was divided by 2 and 4. Setting a at 16 in Eq. (2) gives predictions corresponding to the lower dashed curve in Fig. 2.

It was *a priori* obvious that Horst's model could not account for all of our data since this model predicted identical thresholds in the ''peak'' and ''trough'' conditions. Yet, it could be expected that the model would yield correct predictions in the ''peak'' condition, because its basic assumption was that subjects perform the discrimination task by making measurements of F_{vertex}: This assumption seemed to be valid for ''peak'' stimuli, whatever η. In fact, the model largely failed to predict the effect of η on thresholds for ''peak'' stimuli; paradoxically, the model was rather more successful for ''trough'' stimuli. A comparison of the results obtained in the ''peak'' condition with the lower dashed curve (giving, in principle, more reasonable predictions than the upper dashed curve) indicates that the measured thresholds were higher than predicted for $\eta>3$, but *lower* than predicted for the smallest values of η. For $\eta>3$, it could be expected that the predictions would be too low, and that the error would increase with η, because: (1) according to the model, the width of the temporal window used by the subjects exceeded 25 ms (see Table I); (2) for durations exceeding about 25 ms, the ''uncertainty principle'' reflected by the last term of Eq. (2) is no longer verified psychophysically (Chih-an and Chistovich, 1960; Moore, 1973). When η did not exceed 3, however, the temporal windows minimizing S were always short enough to justify the last term of Eq. (2). Thus, for $\eta\leqslant 3$, the discrepancy between the predicted and measured effects of η is remarkable.

Actually, a related discrepancy is apparent in Horst's paper. In his experiment, he varied the depth of the standard FM while keeping constant the standard value of F_{vertex} (always a peak). His model predicted that thresholds would increase with the depth of the standard FM. However, as long as this parameter was larger than 0, the measured thresholds were approximately constant. This discrepancy, and the one found here, suggest that one cause of the peak/trough effect is some auditory mechanism permitting an ''abnormally accurate'' perception of FM peaks. One of the experiments reported in paper 3 suggested instead that the perception of FM troughs is impaired by a deleterious factor. It is of course possible that one and the same mechanism, unidentified as yet, has a beneficial effect on the perception of peaks *and* a deleterious effect on the perception of troughs.

II. EXPERIMENT 2

A. Rationale and method

In experiment 1, as well as those reported in papers 1–3, we were concerned with the perception of frequency extrema that are both preceded and followed by a frequency glide. An obvious question to ask about the perceptual asymmetry found in these experiments is then: Does it originate mainly from a ''forward'' effect of the preceding glide's direction, or on the contrary from a ''backward'' effect of the following glide's direction? It is rather easy to imagine how a perceptual advantage of FM peaks over FM troughs could be produced by a forward effect involving neural inhibition or neural facilitation. The necessary physiological ingredients can be found, for instance, in the cochlear nucleus and the inferior colliculus of the cat (see, e.g., Ehret, 1992). So, a mechanism acting forward might exist below the auditory cortex. On the other hand, a mechanism acting backward would probably be located in the auditory cortex itself, as the effect should be understood as a kind of mnemonic interference or ''recognition masking'' (Massaro and Idson, 1977). In studies using rapid sequences of tone bursts, Watson *et al.* (1975; see also Divenyi and Hirsh, 1975) found that the perceptual encoding of the frequency of one tone can be markedly impaired by the subsequent presentation of *higher-frequency* tones. It is tempting to think that a similar frequency-asymmetric backward interference effect plays a role in FM perception (even though stimulus uncertainty appears to have opposite effects on the perceptual asymmetry of FM and the phenomenon observed by Watson *et al.*, as shown in paper 2).

A simple way to determine the relative weights of forward and backward effects in the perceptual asymmetry of FM is to use *unidirectional* FMs, i.e., constantly rising or constantly falling frequency glides, and to measure discrimination thresholds for the frequency extremum reached at each temporal extremity. A source of perceptual asymmetry

acting forward is demonstrated if thresholds are lower at the end of a rising glide than at the end of a falling glide. A source acting backward is demonstrated if thresholds are lower at the start of a falling glide than at the start of a rising glide. For fair comparisons, the standard frequency extremum should be the same in the four experimental conditions (2 directions$\times$2 temporal extremities).

Some recent studies on the discrimination of stimuli imitating speech formant transitions (Porter *et al.*, 1991; Collins *et al.*, 1990; van Wieringen and Pols, 1995) suggested that a perceptual advantage of FM peaks over FM troughs can be produced both forward and backward. The study by Porter *et al.* was already described in paper 1. For the *starting* point of formant transitions lasting 60 or 120 ms, these authors measured lower thresholds with falling transitions than with rising transitions. However, their standard rising transitions started at 1500 Hz, whereas the standard falling transitions started at 2100 Hz, and they ascribed the perceptual advantage of falling transitions to this difference in frequency register rather than to an intrinsic advantage of frequency peaks over frequency troughs. For the discrimination between the *ends* of formant transitions, Collins *et al.* (1990) found, on the opposite, a perceptual advantage of *rising* transitions over falling transitions; but the factor of transition direction was again confounded with a frequency register factor. In experiments similar to those of Porter *et al.* and Collins *et al.*, van Wieringen and Pols (1995) obtained results from which they concluded that there was no significant effect of transition direction. However, this negative conclusion appears to be based on the fact that their discrimination thresholds were measured in Hz, that is in terms of absolute frequency differences: If the data displayed in their Table III are converted into *relative* frequency differences, the logical conclusion becomes that frequency peaks have a marked perceptual advantage over frequency troughs, both at the beginning and at the end of formant transitions.

The unidirectional FMs used in experiment 2 had the shape of one-half of a cosine cycle on a log frequency scale. They took place in the central 100 ms of stimuli which had a total duration of 110 ms and were gated on and off with 10-ms amplitude ramps. The amplitude ramps were cosinusoidal (same shape as the FMs). During the first 5 and last 5 milliseconds of the stimuli, f_{inst} was kept on a plateau corresponding to the extremum reached by the FM.[1] In the standard stimuli, the critical frequency extremum was always at 1000 Hz. The standard span of f_{inst} was equal to 0, 600, or 1800 cents. When this standard span was 600 cents, the stimuli could be described as the first-half or second-half of stimuli that we previously employed (see especially paper 2). For each type of critical frequency extremum (end of rise, end of fall, start of rise, or start of fall), Fig. 3 depicts the standard stimuli (thick lines) and target stimuli (thin lines) corresponding to a standard span of 0 cent (left part of the panels) and a larger standard span (right part of the panels). On each trial, as before, the span of f_{inst} was larger in the target stimulus than in the standard stimulus.

Six listeners with normal hearing took part in the experiment. Two of them (AS and YB) were psychoacoustically naive at the outset. For these subjects, in addition to sounds

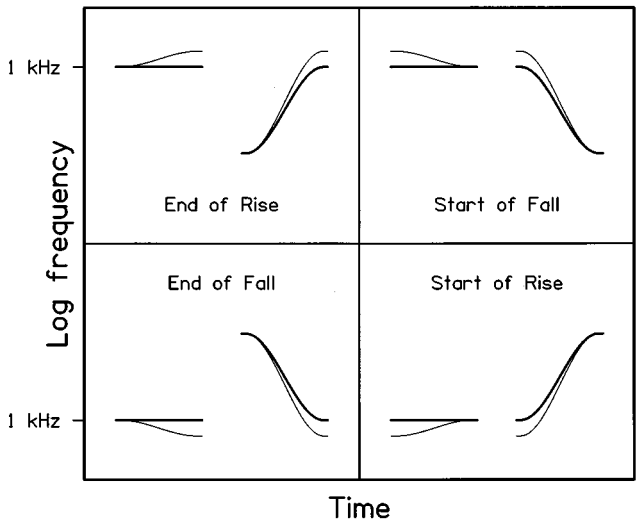


FIG. 3. Schematic spectrograms of the stimuli used in the four main conditions of experiment 2. Standard and target stimuli are, respectively, represented by thick and thin lines. The standard stimuli were either steady tones or frequency modulated tones, as shown, respectively, in the left and right parts of each panel.

containing a unidirectional FM, the stimulus set included 200-ms ''control'' stimuli in which the FM consisted of one *full* cycle of a 5-Hz cosine function and the standard span of f_{inst} was 600 cents. With these stimuli, as in experiment 1, we measured discrimination thresholds for the temporally central frequency vertex (standard value: 1000 Hz) in a ''control peak'' condition and a ''control trough'' condition. For the other four subjects, including the two authors, such data had already been collected in other studies.

All the stimuli were binaurally presented at 70 dB SPL.[2] They were generated as in experiment 1 (except that f_{inst} was now sampled at an even higher rate). The adaptive procedure used to measure thresholds was the same as before. In the conditions involving unidirectional FMs, only three and two subjects were, respectively, tested with standard spans of 0 and 1800 cents; for a given subject, extremum type, and standard span, at least 20 threshold measurements were made. For both unidirectional and bidirectional FMs, the thresholds reported below are (arithmetic) means of the last ten measurements made in each condition.

B. Results and discussion

Table II presents the thresholds obtained from each subject in each condition. The standard errors of the measurements (in parentheses) amount to about 10% of the means. In order to quantify the peak/trough asymmetries which could be ascribed to (1) a purely ''forward'' action, (2) a purely ''backward'' action, (3) the sum of (1) and (2), we computed the ratios of the thresholds obtained in the conditions: (1) ''end of rise'' and ''end of fall'' (EF/ER); (2) ''start of fall'' and ''start of rise'' (SR/SF); (3) ''control peak'' and ''control trough'' (CT/CP). The obtained ratios are displayed in Fig. 4.

Let us consider first the ratios obtained when the standard span of f_{inst} was 600 cents (middle panel of Fig. 4). The geometric means of the EF/ER and SR/SF ratios were, re-

TABLE II. Discrimination thresholds measured in experiments 2 and 3. These thresholds are expressed in cents. The standard errors of the threshold measures (10 measures for each cell) are given in parentheses. The means across subjects are geometric. Numbers in italics represent data collected in the course of other studies.

Subject	End of rise	End of fall	Start of fall	Start of rise	Control peak	Control trough
			Experiment 2			
Standard span of f_{inst}: 0 cent						
SC	7.6 (1.0)	6.5 (0.7)	17.1 (2.8)	23.5 (1.1)		
LD	4.3 (0.6)	7.4 (0.9)	11.7 (1.1)	18.5 (1.8)		
JJ	14.2 (3.2)	17.9 (1.5)	28.2 (3.1)	26.3 (2.4)		
Mean	7.7	9.5	17.8	22.5		
Standard span of f_{inst}: 600 cents						
SC	10.5 (1.1)	23.4 (3.4)	16.1 (1.2)	93.0 (9.5)	*9.4 (0.9)*	*81.3 (7.0)*
LD	7.1 (0.6)	24.9 (1.6)	24.2 (1.8)	53.0 (6.5)	*6.3 (1.1)*	*28.0 (2.2)*
JJ	22.8 (2.1)	29.2 (3.5)	44.7 (4.8)	43.3 (3.7)	*11.6 (1.2)*	*32.9 (3.6)*
AS	12.5 (1.1)	19.6 (3.2)	88.7 (8.2)	55.3 (4.1)	14.6 (1.3)	77.2 (8.9)
MP	11.8 (1.5)	47.2 (4.4)	46.5 (5.0)	103.7 (10.6)	*12.4 (0.8)*	*64.7 (4.7)*
YB	16.2 (1.5)	22.2 (2.2)	52.3 (6.7)	63.2 (6.9)	23.8 (2.3)	53.8 (4.3)
Mean	12.6	26.5	39.4	65.3	12.0	52.2
Standard span of f_{inst}: 1800 cents						
SC	11.2 (1.3)	24.8 (2.4)	18.6 (1.6)	76.8 (7.0)		
LD	11.1 (1.5)	34.8 (4.7)	35.8 (6.2)	48.1 (5.3)		
Mean	11.1	29.4	25.8	60.8		
			Experiment 3			
Standard span of f_{inst}: 600 cents						
LD	17.2 (1.4)	58.2 (6.2)	17.8 (2.3)	113.8 (8.2)	5.5 (0.6)	25.9 (1.5)
JJ	43.7 (1.5)	88.4 (6.9)	48.0 (3.2)	81.8 (6.0)	11.6 (1.2)	32.9 (3.6)
JL	25.0 (1.7)	65.3 (5.4)	24.1 (2.8)	59.3 (5.6)	9.7 (1.0)	29.2 (2.6)
TV	70.7 (5.3)	203.1 (15.7)	46.6 (4.0)	154.8 (22.2)	20.1 (2.8)	49.7 (5.6)
Mean	34.0	90.9	31.3	96.2	10.6	33.3

spectively, 2.10 and 1.65. These two values are larger than 1 and rather similar, which suggests that the peak/trough effect can be produced both forward and backward, with approximately equal strengths. However, looking at the individual data, one notes that EF/ER was larger than SR/SF in five subjects out of six and that the SR/SF ratios were not always larger than 1. Thus, the evidence for a retroactive creation of the peak/trough effect is not spectacular. Overall, both the SR/SF and EF/ER ratios were rather small. For each subject, the CT/CP ratio was larger. The geometric mean of its individual values was 4.35.

When the standard span of f_{inst} was 0 (upper panel of Fig. 4), there was no reason to expect that the EF/ER and SR/SF ratios would differ markedly from 1; they were indeed close to 1. On the other hand, when the standard span of f_{inst} was 1800 cents (lower panel), ratios larger than 1 were again expected, and obtained. It is interesting to consider, for the two subjects who were tested at each standard span (SC and LD), the effect of this variable on the measured thresholds. For each extremum type, as shown in Table II, thresholds increased notably when the standard span varied from 0 to 600 cents (except for subject SC in the SF condition). However, a further widening of the standard span, up to 1800 cents, had little effect. This pattern of results is similar to that observed by Horst (1989) for sounds resembling our CP stimuli (cf. Sec. I D of the present paper).

Table II also shows that for each standard span, frequency extrema in final position were better discriminated than frequency extrema in initial position. Similar end/start asymmetries in frequency discrimination were previously reported in a number of papers concerning mostly the perception of formant transitions (Mattingly *et al.*, 1971; Collins, 1984; Cosgrove *et al.*, 1989; Elliott *et al.*, 1992; van Wieringen and Pols, 1995). This phenomenon is probably related to some observations on the pitch of stimuli consisting of short frequency glides: Brady *et al.* (1961) and Nábělek *et al.* (1970) found that the *final* frequency content of such stimuli is very heavily weighted by the pitch extractor. It was also found by the same authors that this ''pitch bias'' is stronger for rising glides than for falling glides. The latter finding is consistent with the fact that we obtained lower thresholds in the ER condition than in the EF condition.

To some extent, the CP condition could be considered as a mere juxtaposition of the ER and SF conditions (for a standard span of 600 cents). Thus a crude model of discrimination in the CP condition predicted that, for each subject, the CP threshold would be equal to either the ER threshold or the SF threshold, whichever was the lowest (and it was always the ER threshold). Similarly, it could be predicted that the CT threshold would be equal to the EF threshold, which was always lower than the SR threshold. An examination of Table II reveals that the first of these two predictions worked well; its only serious error was an overestimation of JJ's CP threshold by a factor of 2. However, the second prediction did not work at all for three of the six subjects (SC, AS, and YB), who performed much worse than predicted in the CT condition. The success of the first prediction, as well as the failure of the second one, tally with

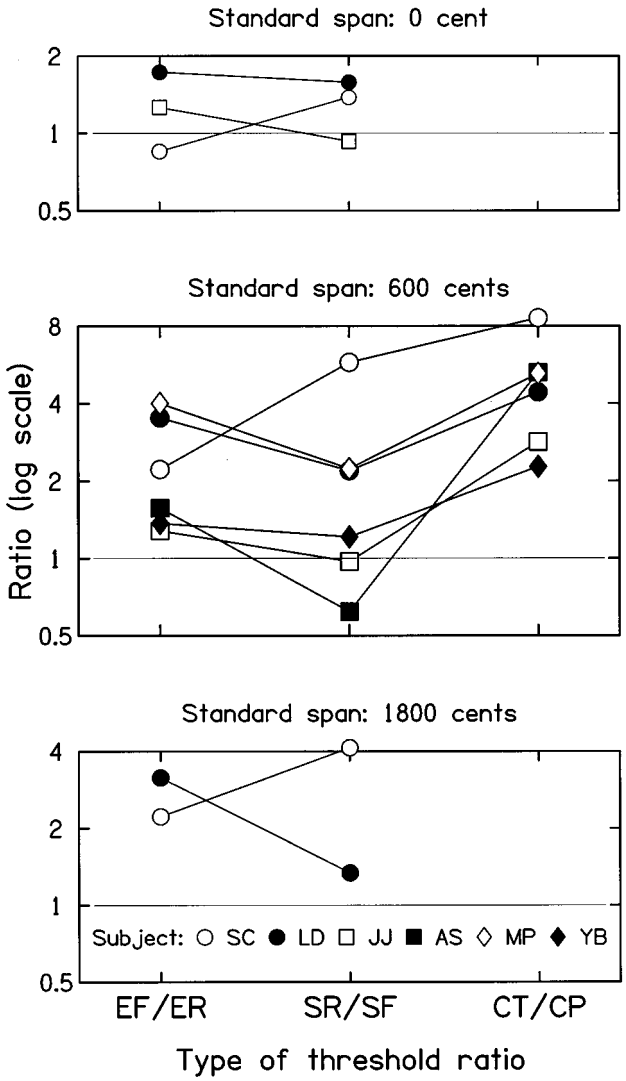


FIG. 4. Threshold ratios computed from the data displayed in the upper part of Table II (results of experiment 2). Each subject is represented by a specific symbol. Ratios larger than 1 reflect a perceptual advantage of frequency maxima over frequency minima. This perceptual advantage could be produced by the previous acoustic context (EF/ER), the following acoustic context (SR/SF), or both contexts (CT/CP).

subjects' conscious impressions about the stimuli: In the CP stimuli, the initial rising glide is perceptually much more salient than the following falling glide; introspectively, therefore, the CP condition was similar to the ER condition; by contrast, the CT stimuli did not sound similar to either purely rising or purely falling glides.

III. EXPERIMENT 3

A. Rationale and method

In experiment 2, we obtained lower thresholds when the critical frequency extremum was at the final point of the stimuli than when it was at the starting point. For this reason, comparing the EF/ER and SR/SF ratios was a somewhat problematic way to assess the relative weights of ''forward'' and ''backward'' actions in the production of a perceptual asymmetry between FM peaks and troughs. In experiment 3, we wished to assess again the relative weights of forward

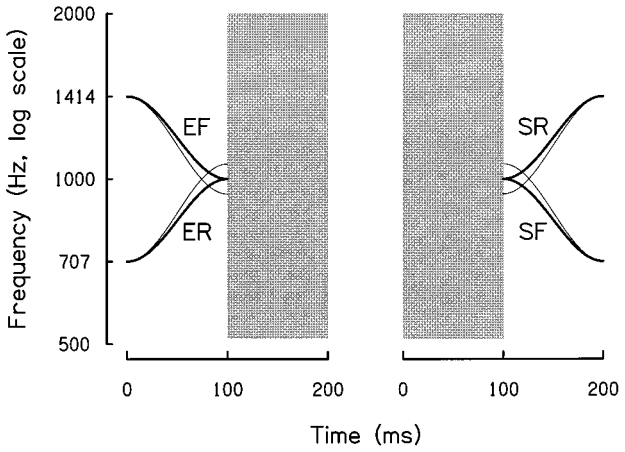


FIG. 5. Schematic spectrograms of the stimuli used in the four main conditions of experiment 3. Standard and target stimuli are, respectively, represented by thick and thin lines. The filled rectangles represent noise bursts.

and backward actions, but using stimuli in which the critical frequency extremum would always have a temporally central position, as in most of our previous experiments. The solution was to re-employ the 200-ms CP and CT stimuli of experiment 2 and to mix them with a 100-ms noise burst coinciding with, and masking, either their second-half (for the assessment of forward actions) or their first-half (for the assessment of backward actions). In doing so, as shown in Fig. 5, we produced four experimental conditions which were labeled again as ''end of rise'' (ER), ''end of fall'' (EF), ''start of rise'' (SR), and ''start of fall'' (SF). Essentially, these four conditions differed from the corresponding conditions of experiment 2 only in that the critical frequency extrema were adjacent to a noise (with a large bandwidth) instead of a silence. It was especially interesting to compare, after this addition of noise, the SR and SF thresholds. In the absence of noise, when the critical frequency extrema of conditions SR and SF coincided with a stimulus onset, the neural encoding of these frequency extrema was probably affected by ''adaptation'' phenomena (see, e.g., Palmer, 1995). By contrast, when we introduced noise just before a critical frequency extremum, the neural response to this frequency extremum could be considered as a ''post-adaptation'' response. In each of the four experimental conditions, frequency discrimination was of course likely to be impaired by the noise bursts preceding or following the critical extremum. However, this was not, *per se*, a methodological weakness. The essential point was that, due to their physical properties (see below), the noise bursts were not liable to produce a larger forward masking effect in condition SR than in condition SF, or a larger backward masking effect in condition EF than in condition ER. More generally, there was no reason to think that the noise bursts could give an artificial advantage to the perception of frequency maxima.

The standard span of f_{inst} was fixed at 600 cents. As in experiment 2, the standard value of each critical frequency extremum was 1000 Hz. However, the (partially masked) CP and CT stimuli were presented at 55 instead of 70 dB. We did so in order to avoid having to present the concurrent noise bursts at a high SPL. The noise bursts were obtained

by bandpass filtering the output of a pink noise source between 500 and 2000 Hz. They were gated on and off abruptly. The rejection slopes of the bandpass filter (Kemo VBF 8) were about 75 dB/oct. For each subject, the levels of the noise just necessary to mask the standard CP stimulus, and the standard CT stimulus, were determined in a 2I-2AFC paradigm with an adaptive procedure which was similar to that used for the measurement of the frequency discrimination thresholds.[3] In each trial, the two listening intervals contained a 200-ms noise burst. The two noise bursts were independent of each other but had the same SPL. In one of the intervals, to be identified by the subject, the noise was mixed with the 55-dB CP or CT stimulus. At the outset of a trial block, the noise SPL was always small enough to make the subject's task easy. Following a correct response, the noise SPL was increased by 3 dB (initially) or 1.5 dB (after the second reversal). Following a wrong response, the noise SPL was decreased by 9 dB (initially) or 4.5 dB (after the second reversal). The SPL for which 75% of responses were correct was estimated as the arithmetic mean of the SPLs used on the 12 trials corresponding to reversals 3–14. For each subject, more than 20 measurements of this ''threshold SPL'' were made with the CP stimulus and (in separate blocks of trials) the CT stimulus. The grand means of the measured threshold SPLs were 64.1 dB for the CP stimulus and 65.8 dB for the CT stimulus. A slight advantage for the CT stimulus was observed in each subject. In the main part of the experiment, the SPL of the 100-ms noise bursts was set at the threshold SPL for the CT stimulus; it was fixed across conditions, but varied slightly across subjects, in a 2-dB range.

In conditions ER, EF, SF, and SR, at least 30 measurements of the frequency discrimination threshold were made for a given subject. The mean threshold values reported in the next section were computed from the last ten measurements. Frequency discrimination was also assessed in the absence of noise, in a ''CP'' and a ''CT'' condition. These two conditions replicated those labeled identically in experiment 2, at 55 instead of 70 dB SPL. In each of them, again, the formal data were collected only after several practice sessions and consisted of ten threshold measurements.

The experiment was performed on four listeners with normal hearing. Two of them (LD and JJ) were also subjects in experiment 2. The other two listeners had no previous experience with psychoacoustic tasks.

B. Results and discussion

The results are displayed in the lower part of Table II and in Fig. 6. They were simpler and more clear-cut than those obtained in experiment 2 for the same standard span of f_{inst} (600 cents). Looking first at the EF/ER and SR/SF ratios (Fig. 6), one notes that they were globally somewhat larger than in experiment 2, and less variable from subject to subject. The mean values of EF/ER (2.67) and SR/SF (3.07) were similar to each other, and also similar to the mean value of CT/CP (3.16). Another important fact was that, this time, very similar thresholds were obtained in conditions ER and SF, as well as in conditions EF and SR (see Table II). Thus, in the four main experimental conditions, thresholds did not

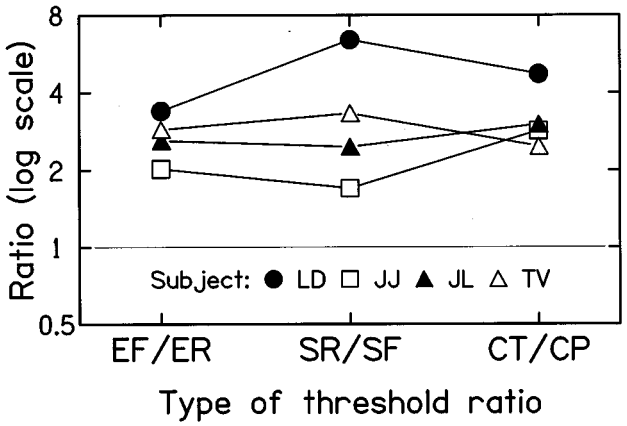


FIG. 6. Threshold ratios computed from the data displayed in the lower part of Table II (results of experiment 3).

depend on whether the critical frequency extremum was at the starting point or the final point of the audible FM. What only mattered was the nature of the extremum in the *frequency* domain, i.e., whether it was a maximum frequency or a minimum frequency. Clearly, these results provide strong evidence for both forward-acting and backward-acting sources of the peak/trough effect.

Note that the mean thresholds measured in conditions EF and SR were very large (90.9 and 96.2 cents). The corresponding frequency shifts produced a change in FM slope of about 15%. Apparently, smaller changes in FM slope could not be detected and used to perform the task. This is rather surprising in view of the suggestion by Dooley and Moore (1988) that listeners can detect 6% changes in a FM slope. However, the stimuli used by Dooley and Moore differed in many respects from those used in the present experiment, so that there is no flagrant discrepancy between the two sets of results.

From the finding that the peak/trough effect can be produced both forward and backward, with approximately equal strengths, what do we learn about the *mechanisms* of this perceptual phenomenon and their location in the auditory system? The existence of retroactive factors argues against the idea that the peak/trough effect would completely originate from cochlear mechanisms and thus be reflected in some aspect of the eighth nerve's responses to FM stimuli (see, in this respect, Duifhuis, 1973). In paper 3 (Sec. III), we hypothesized that the peak/trough effect reflects properties specific to a *temporal* coding of instantaneous frequency by the auditory system. The present data do not support this hypothesis, as they suggest that at least part of the peak/trough effect has a central origin: Due to neural phase-locking, there is a temporal representation of instantaneous frequency in the auditory nerve (Sinex and Geisler, 1981); beyond the cochlear nuclei, however, it seems that neural phase-locking rapidly disappears (de Ribaupierre *et al.*, 1972; Steinschneider *et al.*, 1980).

To account for a retroactive production of the peak/trough effect, it seems necessary to admit that within the neural network that *stores* auditory information concerning frequency extrema—a network probably located in the auditory cortex—the representation of a local frequency mini-

mum tends to be degraded by a subsequent frequency rise. The production of the peak/trough effect in a *forward* direction might rest on completely different mechanisms, unrelated to auditory memory *per se* and located subcortically. Note, however, that the cochlear nonlinearity termed as ''two-tone suppression'' cannot play a crucial role since this process does not significantly persist after the cessation of the suppressor (Arthur *et al.*, 1971; Stanny and Elfner, 1980). A parsimonious hypothesis is that the neural interactions producing the peak/trough effect forward are actually the same as those acting backward. Clément (1996) designed a formal neural network intended to model crudely the perception of instantaneous frequency. The network consists of two layers of tonotopically organized cells with asymmetric inhibitory connections (in the first layer) and temporal integration properties. When its input is a continuous FM, the network essentially identifies the local frequency maxima. It is less sensitive to local frequency minima, and what we wish to stress here is that this asymmetry occurs both at the starting point and the final point of stimuli consisting of a unidirectional glide. In the model, therefore, the peak/trough effect is produced both forward and backward, by identical (and physiologically plausible) mechanisms. It remains to be seen, among other things, if the model correctly simulates the influence of glide duration. The results of Porter *et al.* (1991), as well as pilot data collected in our laboratory, suggest that the peak/trough effect is no longer observed at very short stimulus durations. Further studies on this issue are needed.

IV. CONCLUSIONS

Overall, the present experiments show that the perceptual asymmetry between FM peaks and troughs (the ''peak/trough effect'') does not have a simple origin. In experiment 1, using a family of modulation waveforms ranging from quasi-triangles to quasi-squares, we found that the peak/trough effect was maximum for the most triangular modulations, with essentially durationless vertices. The detection threshold of shifts in a durationless frequency maximum appeared to be surprisingly small (about 10 cents, i.e., 0.6%). This suggests that some unknown auditory mechanism permits a *hyperacute* perception of frequency maxima, and thus that the peak/trough effect is not reducible to the effect of a deleterious factor on the perception of frequency minima. By showing that the peak/trough effect can be produced *retroactively*, experiments 2 and 3 made clear that its origin is at least partly central (certainly post-cochlear and maybe cortical). This does not support our previous hypothesis that the peak/trough effect reflects properties specific to a *temporal* coding of instantaneous frequency by the auditory system.

ACKNOWLEDGMENTS

We thank Alain de Cheveigné, J. Wiebe Horst, and John P. Madden for comments on a previous version of this paper. We are also grateful to Jean-Jacques Sierra for his active participation in the experiments.

[1]Thus the amplitude ramps were twice as long as the frequency plateaux. Another methodological option was to use equally long amplitude ramps and frequency plateaux. However, the frequency plateaux had to be short, for obvious reasons, whereas the amplitude ramps had to be relatively long in order to minimize spectral splatter.

[2]For subject JJ, in the ''control peak'' and ''control trough'' conditions, the SPL was 55 dB instead of 70 dB. The corresponding data were actually collected during experiment 3.

[3]Nábělek (1978; see also Collins and Cullen, 1984) measured the detection thresholds, in a wideband masking noise, of gliding tones that were comparable to the first-half and the second-half of our CP and CT stimuli. For tone durations of 50–200 ms, thresholds appeared to be independent of glide direction. Here, therefore, we could assume that when the noise just masked an entire CP or CT stimulus, both the first-half and the second-half of this stimulus were just masked. This is why we did not measure masking thresholds for the two halves separately.

Arthur, R. M., Pfeiffer, R. R., and Suga, N. (**1971**). ''Properties of 'two-tone inhibition' in primary auditory neurones,'' J. Physiol. (London) **212**, 593–609.

Brady, P. T., House, A. S., and Stevens, K. N. (**1961**). ''Perception of sounds characterized by a rapidly changing resonant frequency,'' J. Acoust. Soc. Am. **33**, 1357–1362.

Chih-an, L., and Chistovich, L. A. (**1960**). ''Frequency-difference limens as a function of tonal duration,'' Sov. Phys. Acoust. **6**, 75–80.

Clément, S. (**1996**). ''Modélisation de la perception auditive des modulations de fréquence,''Master's thesis in Cognitive Science, Université Bordeaux 2, France.

Collins, M. J. (**1984**). ''Tone glide discrimination: Normal and hearing-impaired listeners,'' J. Speech Hear. Res. **27**, 403–412.

Collins, M. J., and Cullen, J. K. (**1984**). ''Effects of background noise level on detection of tone glides,'' J. Acoust. Soc. Am. **76**, 1696–1698.

Collins, M. J., Cullen, J. K., Jackson, D. F., and Porter, R. J. (**1990**). ''Discrimination of formant-like frequency transitions preceded or followed by steady states,'' Abstracts of the Thirteenth Midwinter Meeting of the Association for Research in Otolaryngology, pp. 181–182.

Cosgrove, P., Wilson, J. P., and Patterson, R. D. (**1989**). ''Formant transition detection in isolated vowels with transitions in initial and final position,'' Proc. IEEE ICASSP **1**, 278–281 (Bell & Brian, Glasgow).

Demany, L., and McAnally, K. I. (**1994**). ''The perception of frequency peaks and troughs in wide frequency modulations,'' J. Acoust. Soc. Am. **96**, 706–715.

Demany, L., and Clément, S. (**1995a**). ''The perception of frequency peaks and troughs in wide frequency modulations. II. Effects of frequency register, stimulus uncertainty, and intensity,'' J. Acoust. Soc. Am. **97**, 2454–2459.

Demany, L., and Clément, S. (**1995b**). ''The perception of frequency peaks and troughs in wide frequency modulations. III. Complex carriers,'' J. Acoust. Soc. Am. **98**, 2515–2523.

Divenyi, P. L., and Hirsh, I. J. (**1975**). ''The effect of blanking on the identification of temporal order in three-tone sequences,'' Percept. Psychophys. **17**, 246–252.

Dooley, G. J., and Moore, B. C. J. (**1988**). ''Duration discrimination of steady and gliding tones: A new method for estimating sensitivity to rate of change,'' J. Acoust. Soc. Am. **84**, 1332–1337.

Duifhuis, H. (**1973**). ''Consequences of peripheral frequency selectivity for nonsimultaneous masking,'' J. Acoust. Soc. Am. **54**, 1471–1488.

Ehret, G. (**1992**). ''Le mésencéphale auditif, une 'gare de triage' du traitement de l'information acoustique,'' in *Le Système Auditif Central*, edited by R. Romand (INSERM, Paris).

Elliott, L. L., Hammer, M. A., and Carrell, T. (**1992**). ''Discrimination of converging and diverging frequency transition,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford).

Hartmann, W. M. (**1997**). *Signals, Sound, and Sensation* (AIP, Woodbury, NY).

Horst, J. W. (**1989**). ''Detection and discrimination of frequency modulation of complex signals,'' J. Acoust. Soc. Am. **85**, 2022–2030.

Massaro, D. W., and Idson, W. L. (**1977**). ''Backward recognition masking in relative pitch judgments,'' Percept. Mot. Skills **45**, 87–97.

Mattingly, I. G., Liberman, A. M., Syrdal, A. K., and Halwes, T. (**1971**). ''Discrimination in speech and nonspeech modes,'' Cogn. Psychol. **2**, 131–157.

Moore, B. C. J. (**1973**). ''Frequency difference limens for short-duration tones,'' J. Acoust. Soc. Am. **54**, 610–619.

Náběl ek, I. V. (**1978**). ''Temporal summation of constant and gliding tones at masked auditory threshold,'' J. Acoust. Soc. Am. **64**, 751–763.
Náběl ek, I. V., Náběl ek, A. K., and Hirsh, I. J. (**1970**). ''Pitch of tone bursts of changing frequency,'' J. Acoust. Soc. Am. **48**, 536–553.
Palmer, A. R. (**1995**). ''Neural signal processing,'' in *Hearing*, edited by B. C. J. Moore (Academic, New York).
Porter, R. J., Cullen, J. K., Collins, M. J., and Jackson, D. F. (**1991**). ''Discrimination of formant transition onset frequency: Psychoacoustic cues at short, moderate, and long durations,'' J. Acoust. Soc. Am. **90**, 1298–1308.
Ribaupierre, F. de, Goldstein, Jr., M. H., and Yeni-Komshian, G. (**1972**). ''Cortical coding of repetitive acoustic pulses,'' Brain Res. **48**, 205–225.
Sinex, D. G., and Geisler, C. D. (**1981**). ''Auditory-nerve fiber responses to frequency-modulated tones,'' Hearing Res. **4**, 127–148.
Stanny, R. R., and Elfner, L. F. (**1980**). ''An 'inhibitory' influence on brainstem population responses,'' Science **208**, 418–419.
Steinschneider, M., Arezzo, J., and Vaughan, Jr., H. G. (**1980**). ''Phase-locked cortical responses to a human speech sound and low-frequency tones in the monkey,'' Brain Res. **198**, 75–84.
Watson, C. S., Wroton, H. W., Kelly, W. J., and Benbassat, C. A. (**1975**). ''Factors in the discrimination of tonal patterns. I. Component frequency, temporal position, and silent intervals,'' J. Acoust. Soc. Am. **57**, 1175–1185.
Wieringen, A. van, and Pols, L. C. W. (**1995**). ''Discrimination of single and complex consonant–vowel- and vowel–consonant-like formant transitions,'' J. Acoust. Soc. Am. **98**, 1304–1312.

Organization and discrimination of repeating sound sequences by newborn infants[a)]

Stephen McAdams[b)]
Laboratoire de Psychologie Expérimentale (CNRS), Université René Descartes, EPHE, 28 rue Serpente, F-75006 Paris, France and Institut de Recherche et de Coordination Acoustique/Musique (IRCAM), 1 place Igor-Stravinsky, F-75004 Paris, France

Josiane Bertoncini
Laboratoire de Sciences Cognitives et Psycholinguistique (CNRS), Ecole des Hautes Etudes en Sciences Sociales, 54 bd Raspail, F-76270 Paris 06, France

(Received 18 September 1996; accepted for publication 28 July 1997)

A study was conducted to determine whether newborn infants organize auditory streams in a manner similar to that of adults. A series of three experiments investigated the ability of 3- to 4-day-old infants to discriminate repeated rising and falling four-tone sequences in two configurations of source timbre and spatial position. It was hypothesized that if the sequences were organized into two auditory streams on the basis of timbre and spatial position, one of the configurations should be discriminable from its reversal while the other should not. The sequences were tested with different pitch and temporal intervals separating the tones. Sequences were discriminated for the first configuration by adults at both fast tempo/small interval and slow tempo/large interval combinations, while only the latter was discriminated by newborns as measured with a non-nutritive high-amplitude sucking paradigm. Neither adults nor infants could discriminate the sequence reversals for the second configuration. The results suggest that newborn infants organize auditory streams on the basis of source timbre and/or spatial position. They also suggest that newborns have limits in temporal and/or pitch resolution when discriminating tone sequences. © *1997 Acoustical Society of America.* [S0001-4966(97)03411-5]

PACS numbers: 43.66.Mk, 43.66.Jh, 43.66.Qp [WJ]

INTRODUCTION

The acoustic environment is composed of numerous sources of sound, and one of the important tasks for any animal is to be able to perceive them separately and to perform actions with respect to them. This ability may be considered to involve the building up of a veridical mental representation of the sources of sound present in the environment, a representation that is then used to plan appropriate action. One class of processes that seems to be involved in this kind of perceptual organization is the connecting together through time of individual sound events that are emitted by the same source, a process called sequential auditory organization or auditory streaming (Bregman and Campbell, 1971; McAdams and Bregman, 1979; Bregman, 1990). Several acoustic factors have been shown to play a role in streaming, such as spectral discontinuity (Bregman *et al.*, 1990), intensity discontinuity (van Noorden, 1977), and spatial discontinuity (Hartmann and Johnson, 1991). That is, a sequence composed of sounds that are more or less similar with respect to these three dimensions tends to be heard as a single sound stream, whereas a sequence that alternates between two regions of relatively distant values along one or more of these dimensions tends to be heard as two streams.

For most sound sources in the everyday world, the various sound dimensions give rise to unambiguous organizations into sound objects and allow a clear understanding of their behavior through time. In addition, this organization appears to take place automatically without much conscious or deliberate intervention on the part of the organism. An important question that arises in the face of this evidence is whether the processes of sequential auditory organization are part of our innate perceptual equipment (present at birth or maturing within the first few weeks of life), or whether they are acquired through experience in the acoustic world. From birth, infants have to deal with an acoustic environment in which many simultaneous sources of sound can compete. Moreover, it is from such noisy situations that infants must correctly extract highly relevant acoustic information such as their mother's voice among others speaking her native language. Surprisingly, there is almost no research on infants' capacity to organize their acoustic world in terms of separate sources, or to perceive auditory streams. Auditory development has received much interest (e.g., Trehub and Trainor, 1993; Werner and Rubel, 1992), particularly in terms of the development of speech perception (e.g., Jusczyk, 1997), but very few studies have addressed the problem of how infants perceive complex sounds as coherent units, similarly to what is called object perception in visual cognition (cf. Spelke, 1990; Spelke *et al.*, 1992). Although scarce, the data available seem nonetheless to favor the existence of unlearned, basic mechanisms engaged in streaming very early in life.

[a)]Portions of these results were presented at the 120th Meeting of the Acoustical Society of America, San Diego [McAdams *et al.*, J. Acoust. Soc. Am. Suppl. 1 **88**, S91 (1990)].
[b)]Electronic mail: smc@ircam.fr

Some arguments in this direction have been proposed by Bregman (1990).

The results of two previous studies (Demany, 1982; Fassbender, 1993) are consistent with such an assumption. They have addressed the question directly by testing young infants in auditory conditions that lead adults to perceive sound sequences as organized into two different streams. Demany (1982) demonstrated with a visual fixation procedure that 1.5- to 3-month-old infants organize sound sequences according to spectral proximity (frequency region of pure tones). Fassbender (1993) further demonstrated with a non-nutritive sucking paradigm that 2- to 5.5-month-old infants organize sound sequences on the basis of frequency proximity, intensity similarity, and spectral similarity. From these results it appears that stream segregation processes are operative in the very first weeks of life, at least when streams of fairly simple sounds are separated according to a number of acoustic dimensions known to be effective in promoting segregation in adults. Other properties of sound sources that are powerful cues for detecting and recognizing auditory events in the natural environment, such as correlated multidimensional variation of source timbre and spatial location, have not yet been tested with infants. Thus the present study aimed to extend previous research in two ways. First, we tested *newborns* to verify whether stream segregation is part of the perceptual apparatus that first encounters the acoustic world. And second, but more importantly, we tested their capacity to use the *timbre and spatial position* of naturalistic sounds to perceptually organize sound sequences in terms of two distinct auditory entities. The timbres of the sound sources (vibraphone and trumpet sounds) as well as their spatial locations were chosen to be sure that the differences were easily discriminable by young infants (see Clifton, 1992, for a review on sound localization in infants, and Trehub *et al.,* 1990, concerning processing of timbre). In this study, we do not question infants' ability to distinguish vibraphone from trumpet or sounds coming from the right or left. We assume that this ability is part of the newborn's perceptual repertoire. Our question is whether such differences in timbre and spatial position are automatically used by newborns' to perceive sound sequences as originating from two individual sources.

A melody discrimination paradigm similar to the one used by Demany (1982) was employed to probe stream formation in both newborn infants and adults. A test of infants' ability to discriminate a melody from its retrograde (or reverse order of the tones) was first performed in order to find stimulus conditions under which they could perform the discrimination using a non-nutritive sucking paradigm (experiments 1 and 2). Similar conditions were presented to adults for comparison (experiment 4). Then the discriminable melody patterns were presented under conditions that adults perceive as two separate sound streams organized on the basis of timbre and spatial position (experiment 3 for newborns, experiment 4 for adults). If the latter sequence is organized by newborns on the basis of timbral and spatial similarity, then the sequence discrimination demonstrated in experiments 1 and 2 should fail in experiment 3.

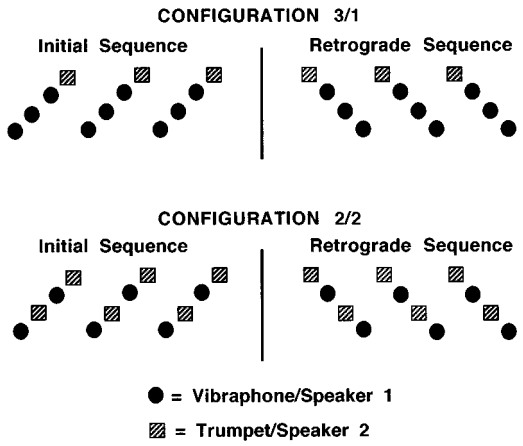


FIG. 1. Three cycles of each of the four-tone repeating stimulus patterns used in the melodic discrimination paradigm are shown. Pitch corresponds to the vertical dimension and time to the horizontal dimension. Timbre and spatial location of the sound sources are shown by the form and shading of the events. Speakers were placed to the right and left of the infant's head.

I. GENERAL METHOD

A. Stimuli

In the interest of giving the infants as many cues as possible for organizing the streams in terms of complex sound sources, it was decided to configure sequences on the basis of two cues, operating in perfect conjunction. Thus two configurations of timbre and spatial configuration were applied to repeating melodic patterns. In these configurations, one speaker was present on each side of the baby's head. Two synthetic timbres created on a Yamaha digital synthesizer were used that simulated a slightly inharmonic, metallic percussion instrument (vibraphone) and a brass wind instrument (trumpet). The instruments that were simulated by the synthesized sound differ in both the resonator (bar versus air column) and the way the resonator is excited (impulsively versus continuously). These specific timbres were previously found by McAdams *et al.* (1995) to be perceived by adult subjects as very dissimilar along several perceptual dimensions (attack time, spectral distribution, and degree of spectral evolution). The timbres were equalized for loudness across the pitch range used in this experiment by author SM. For a given subject, each instrument always appeared in the same speaker and always played the same pitches.

Repeating melodic patterns of four ascending or descending pitches were used (Fig. 1). The experiment was based on the ability to discriminate an initial melodic contour[1] (rising or falling) from its retrograde (falling or rising, respectively). Trehub *et al.* (1987) and Ferland and Mendelson (1989) have demonstrated that 9- to 11-month-old infants can discriminate and categorize these kinds of contours, although, to our knowledge, no data are available on this capacity in newborns.

The first such configuration is labeled 3/1 due to the fact that three tones are played by the vibraphone in one speaker and one tone is played by the trumpet in the other speaker. Under the hypothesis that the sequences would be organized into perceptual streams on the basis of timbre and/or spatial position of the tone events, it was expected that two streams

would arise from these sequences: one with three pitches in a rising or falling pattern, and another with a single repeating pitch. Our experimental hypothesis was that the initial and retrograde patterns would be discriminable due to the difference of the melodic contours in the vibraphone stream. In fact, for this configuration, even if the subject paid no attention to timbre and spatial position, the sequences should be discriminable if the melodic contour has been stored and can be compared across sequences. The reason that two timbres and spatial positions are included in this sequence is to avoid confounding configurational differences between this sequence and the second one to be described below with differences in the number of timbres and spatial positions present in each sequence.

The second configuration is labeled 2/2 since the four pitches are split into two interleaved groups, with pitches 1 and 3 being assigned to the vibraphone and pitches 2 and 4 to the trumpet. Under the hypothesis that the sequences are organized on the basis of these parameters, two perceptual streams should result, each one consisting of a pattern alternating between two pitches. Our experimental hypothesis was that for this configuration, the initial and retrograde patterns would not be discriminable since each stream is perfectly symmetric in terms of its melodic pattern. Therefore, if the newborns can discriminate the 3/1 configuration and cannot discriminate the 2/2 configuration, this result would allow us to argue that they organize sound sequences on the basis of timbre and/or spatial position of the tone events.

Stimuli were synthesized on a Yamaha TX802 FM Tone Generator controlled by a Macintosh computer. The musical instrument simulations were developed by Wessel *et al.* (1987). The stimuli were recorded on a stereo cassette tape.

B. Procedure

An habituation/dishabituation paradigm was used in which sound presentation was controlled by the infant's sucking on a non-nutritive pacifier (Jusczyk, 1985; Floccia *et al.*, 1997). Each time the amplitude of the infant's suck exceeded a fixed threshold, several cycles of a repeating four-tone sequence were played. It was decided to present a minimum of three to four cycles in order to ensure that stream segregation occurred, since Bregman (1978) has shown that this process takes time. A computer program recorded the sucking rate in 1-min periods. The habituation point was considered to have occurred when the sucking rate fell by at least one-third for 2 consecutive min compared to the 1-min period immediately preceding these 2 min. At this point, the stimulus sequence was changed for the experimental group and remained identical for the control group. The infants were considered to have discriminated the sequence change if the difference in mean sucking rate between the 1-min periods before and after habituation was significantly greater for the experimental group than for the control group.

The tapes were played over loudspeakers placed on either side of the infant's head. They were elevated by about 20 cm and formed an azimuthal angle of approximately 120° that was bisected by the orientation of the infant's head. The stimuli were presented at an A-weighted level of about 70 dB in a single-walled sound isolation chamber. The frequency components of the complex sounds used would generally be well above pure-tone thresholds in newborns when presented at this level (cf. Werner, 1992).

Babies were rejected during experimentation if they refused to suck on the pacifier, failed to reach the habituation criterion within 15 min, lost the pacifier during the 3 min prior to or following habituation, fell asleep, cried, or became agitated.

C. Subjects

Subjects in experiments 1–3 were newborn infants of 3 to 4 days of age who were selected to participate in the experiment on the basis of their health record during pregnancy, delivery, and the 3 or 4 days after birth. The selection criteria included the following: (1) their weight at birth had to exceed 2.8 kg; (2) the gestation period had to be at least 38 weeks; (3) their APGAR score (a general measure of health and responsiveness) had to attain the value of 10 at least by the fifth minute following birth; and (4) they had to be in good health at the time of testing. After having obtained the permission of the parents, infants were brought to the experimental situation about 2 1/2 h after their last feeding and 30 min to 1 h before their next one. The newborn infants were tested at the Baudelocque Maternity Hospital in Paris.

II. EXPERIMENT 1

Experiment 1 tested the hypothesis that infants should be able to discriminate the original 3/1 configuration from its retrograde version.

A. Method

Stimulus sequences consisted of a repeating four-note pitch pattern (C4–D4–E4–F♯4, or its retrograde, F♯4–E4–D4–C4) presented at a rate of 10 tones/s (an inter-tone onset interval of 100 ms). Each tone had a duration of 85 ms. Pitches were assigned to either the left or right channel in order to effect spatial separation and to one of two instruments in order to effect a separation based on source timbre (Fig. 1). In these sequences, the three lowest pitches (C, D, E) were assigned to one channel and presented with the vibraphone timbre. The highest pitch (F♯) was presented with the trumpet timbre in the other channel. The side of presentation of the timbres was counterbalanced across subjects. The initial pattern was either the ascending or descending contour, each being presented to an equal number of subjects.

The subjects were randomly assigned to one of two independent groups: Experimental, with contour change at habituation, and Control, with no contour change. Twenty subjects completed the experiment in each group for a total of 40 subjects. Data for 32 additional subjects were rejected: 17 did not habituate within 15 min, 4 produced insufficient or irregular sucking or spat out the pacifier, 3 fell asleep, 7 cried or became agitated, and 1 was removed due to experimenter error.

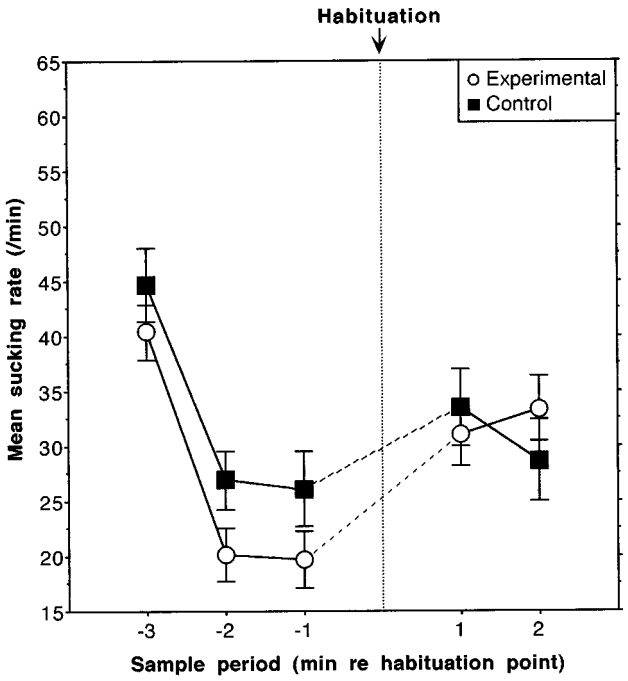


FIG. 2. Results for experiment 1. Mean sucking rate is shown for 1-min sample periods before and after the habituation criterion was attained. Separate curves are shown for experimental (melodic contour change at habituation) and control (no change) groups. Vertical bars represent ± one standard error.

Sound presentation was contingent upon the infant's sucking behavior. Each time a high-amplitude suck was detected by the computer, a ramped gate was opened that allowed the continuous sound sequence on tape to be heard. The ramp lasted half the duration of a four-tone cycle. After the ramp, a minimum of three complete cycles was presented. If no further sucking was detected during this time the sequence was ramped off over half a cycle. Each high-amplitude suck detected during sound presentation resulted in the continuation of sound presentation for 1.2 s following the suck. In general, sucking behavior in newborns occurs in bursts of regular sucking that last several seconds and have a rate of 1.5 to 2 sucks/s. Thus when the infants in our study maintained a rate of at least 0.83 sucks/s during a burst, the sound would be presented continuously during the burst.

B. Results and discussion

The data for three subjects (two control and one experimental) were subsequently rejected since their sucking rate had fallen below seven sucks/min in the 1-min periods either preceding or succeeding the habituation point. This rejection is justified by the fact that a very low sucking rate may give rise first to very few stimulus cycles at a crucial point in the experiment, and second to an exceedingly long silent interval between the temporally adjacent presentations of the pre- and post-habituation stimulus sequences. The average sucking rates for the two groups (18 control, 19 experimental) in the three 1-min periods preceding and the two 1-min periods succeeding the habituation point are shown in Fig. 2. No difference between groups was found in the three 1-min periods prior to habituation according to a repeated measures ANOVA [$F(2,70)<1$], indicating similar behavior during the habituation phase in both groups. It appears that the sucking rate of the experimental group increases slightly more than that of the control group after the habituation point, due to the fact that rate for the experimental group is slightly lower in the pre-habituation period. In order to test the amount of change across the habituation point, a mixed ANOVA was performed with sucking rate as dependent variable and with within-subjects factor Period (1-min periods before and after habituation) and between-subjects factor group (control, experimental). The change in average rate across the habituation point for the control group was not significantly different from that for the experimental group [$F(1,35)<1$]. Identical results are obtained if average sucking rate is computed in 2-min periods preceding and succeeding habituation, although additional subjects must be removed whose sucking rates do not reach criterion in the newly included 1-min periods. These results may be interpreted as indicating that newborn infants cannot discriminate these rising and falling sequences.

In the absence of data on newborns for this kind of discrimination, we chose the values of pitch interval and tempo used in this experiment since they generally give good rising/falling contour discrimination in adults. It is possible that either the sequences were too rapid or the pitch intervals too small (or both) for the infants to be able to acquire differentiable mental representations of these contours. Our goal in this experiment was not to study the respective effects of each of these sequence variables, but to find a melodic pattern for which original and retrograde versions could be discriminated in order to test our main hypothesis concerning stream organization. Therefore, a lower tempo and greater interval size were employed in experiment 2.

III. EXPERIMENT 2

A. Method

The method was identical to that in experiment 1. The stimulus configuration was identical except for doubling the inter-tone onset time to 200 ms and increasing the pitch interval between adjacent tones to five semitones. The pitch sequences used were thus E3–A3–D4–G4 and its retrograde. When the infant sucked on the pacifier, the continuously running sequence was ramped on over 400 ms (half-cycle of the sequence) and stayed on for an additional two cycles before ramping off over 400 ms if an additional suck did not occur. A rate of at least 0.5 sucks/s thus resulted in continuous sound presentation during a burst of sucking. Twenty subjects completed the experiment in each of the experimental and control groups. Data for 44 additional subjects were rejected: 17 did not habituate within 15 min, 13 produced insufficient or irregular sucking or spat out the pacifier, three fell asleep, and 11 cried or became agitated.

B. Results and discussion

The data for four control subjects were subsequently rejected since their sucking rate had fallen below seven sucks/min in the 1-min periods either preceding or succeeding the

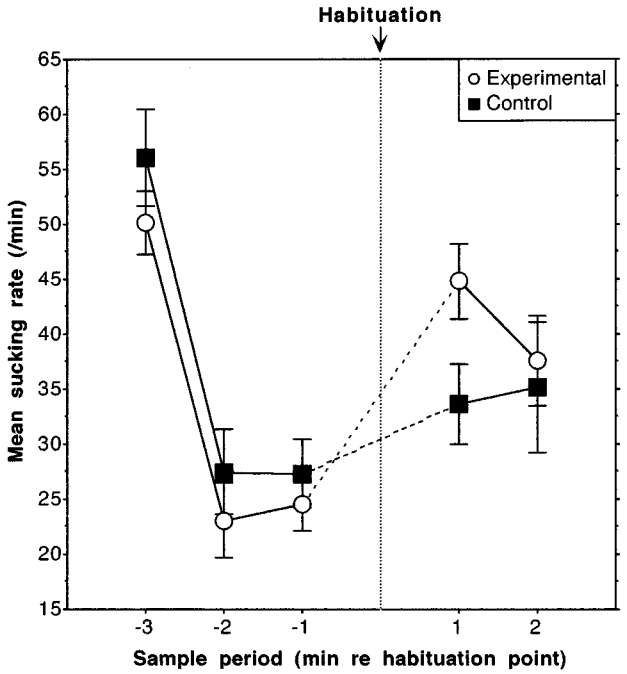


FIG. 3. Results for experiment 2 (see Fig. 2 caption).

habituation point. Data for 20 experimental and 16 control subjects were analyzed. According to a repeated measures ANOVA, no difference between groups was found in the 3-min periods prior to habituation [$F(2,68)<1$](Fig. 3). A Period (2)×Group (2) mixed ANOVA was performed as in experiment 1 and revealed that the experimental group's sucking rate increased more across the habituation point than did that of the control group [$F(1,34)=6.9$, $p<0.05$]. This latter effect is slightly weaker if sucking rate is computed on 2-min periods on either side of the habituation point [$F(1,30)=3.6$, $p=0.066$]. These results indicate that newborn infants can discriminate the rising and falling melodic contours used in this experiment).

Comparisons of the data for experiments 1 and 2 by way of planned contrasts within a between-subjects ANOVA [experiment (2)×group (2)] with the difference in sucking rate between pre-habituation and post-habituation periods as dependent variable clearly show that: (1) across the two experiments performance was similar in the control groups [$F(1,69)<1$] and marginally different between experimental groups [$F(1,69)=3.5$, $p=0.065$, although this difference is nonsignificant if two-minute periods are used to compute mean sucking rate: $F(1,61)=2.4$], and (2) differences between experimental and control groups were not significant for experiment 1 [$F(1,69)<1$] but were significant for experiment 2 [$F(1,69)=8.0$, $p<0.01$]; the same pattern of results was found for sucking rates computed on 2-min periods. This suggests a greater sensitivity to melodic contour change for the slower tempo and larger pitch interval patterns. This difference is unlikely to be related to psychoacoustic limits of frequency discrimination. Frequency resolution is better than or equal to 4% for pure tones at 3 months of age when the sounds are presented at 40 dB SL (Olsho *et al.*, 1987). Frequency resolution is lower at higher frequencies in early infancy and increases over the first year of life, but there is a much smaller effect of development at lower frequencies that are comparable to the fundamental frequencies in our stimuli (Spetner and Olsho, 1990). Further, Olsho (1985) has shown similar psychoacoustic tuning curves in 4-month-olds and adults, although it is unknown whether newborns also have similar tuning to adults.

The better discrimination of slower sequences may also be due to limits in temporal resolution. Gap detection studies by Werner *et al.* (1992) have measured thresholds at about 60 ms in the period of 3–12 months, while thresholds measured in adults are generally less than 10 ms. These thresholds measured in the presence of low-pass noise are up to 100 ms for 3- to 6-month-olds. Corresponding thresholds ($d' \approx 1.0$) in data from Trehub *et al.* (1995) for 6.5-month-old infants were just under 30 ms for tone pips. Again, to our knowledge no data are available on temporal resolution in newborns. The inter-tone intervals in our experiments 1 and 2 were 15 and 115 ms, respectively. Marean and Werner (1991) have shown 20 dB of forward masking of a 1-kHz pure tone by a broadband noise with an inter-tone interval of 20 ms and 12 dB of masking for an interval of 100 ms. While these data are for older infants, the resolution may be even worse at birth. Further research will be needed to tease apart the relative importance of these two factors in the perception of melodic contours by newborn infants and their applicability to discrimination of melodic sequences such as those employed in this study.

IV. EXPERIMENT 3

Having found conditions in which the infants can discriminate the rising and falling sequences, experiment 3 was then run with these same values in order to test the streaming hypothesis. We hypothesized that infants would organize the stimulus sequence into two two-note streams, each of which would not be distinguishable from its retrograde version.

A. Method

The method was identical to that in the previous two experiments. The tempo and pitches of the stimulus sequences were identical to those from experiment 2, but the 2/2 configuration of timbres and spatial positions was used (Fig. 1). Twenty-four subjects completed the experiment in each of the experimental and control groups. Data for 45 additional subjects were rejected: 10 did not habituate within 15 min, 18 produced insufficient or irregular sucking or spat out the pacifier, 5 fell asleep, 10 cried or became agitated, and 2 were removed due to experimenter error.

B. Results and discussion

The data for five subjects (four control and one experimental) were subsequently rejected since their sucking rate had fallen below seven sucks/min in the 1-min periods either preceding or succeeding the habituation point. The data for the two groups (20 control, 23 experimental) are presented in Fig. 4. No difference between groups was found in the 3-min periods prior to habituation [$F(2,82)=1.3$, n.s.]. A mixed ANOVA [period (2)×group (2)] reveals that the experi-

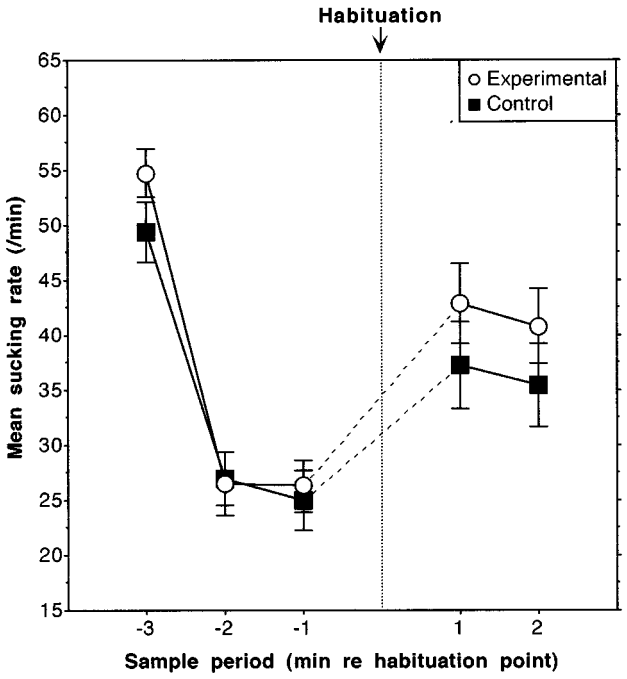


FIG. 4. Results for experiment 3 (see Fig. 2 caption).

mental group's mean sucking rate did not increase more across the habituation point than did that of the control group [$F(1,41)<1$], indicating that they did not discriminate rising and falling contours in the 2/2 configuration. Identical results were found for sucking rates computed on 2-min periods before and after the habituation point.

Data from experiments 2 and 3 were compared by way of planned contrasts within a between-subjects ANOVA [experiment (2)×group (2)] with the difference in sucking rate between pre-habituation and post-habituation periods as dependent variable. Although rates were similar in the control and experimental groups across the two experiments [$F(1,75)<1$ in both cases], differences between experimental and control groups were significant for experiment 2 [$F(1,75)=5.2$, $p<0.05$] but were not significant for experiment 3 [$F(1,75)<1$]. The significant contrast for experiment 2, derived from 1-min sucking periods, is weakened somewhat when 2-min periods are used. This difference may be due to the transient increase in response to novelty of the sequence change for the experimental group as can be seen in Fig. 3. These results suggest that newborns do not discriminate rising from falling patterns when their events are distributed in the 2/2 configuration on the basis of timbre and spatial position.

We can rule out a number of potential explanations for this difference between experiments: (1) The effect cannot be ascribed to general differences in the stimuli since the only difference between experiments 2 and 3 is that one note in the pattern changed timbre and position (transforming the 3/1 configuration into the 2/2 configuration). Otherwise there was no change in stimulus complexity: the melodic patterns had the same pitches, the same intensity, the same tempo, and each had two timbres and two spatial positions. (2) Differences in time elapsed between the last presentation of an habituation stimulus and the first presentation of a new stimulus for the experimental group can also be ruled out. A comparison of critical inter-stimulus intervals across the habituation point show no significant differences between control and experimental groups within each experiment [experiment 2: unpaired $t(34)=-0.09$, n.s.; experiment 3: $t(41)=-0.48$, n.s.] nor differences between corresponding groups across the two experiments [control: $t(34)=1.00$, n.s.; experimental: $t(41)=0.95$, n.s.]. (3) Finally, differences between experiments in newborns' overall response rates are not responsible for the effect either. The planned contrasts described above demonstrate that there was no difference in global response rate between corresponding subject groups in the two experiments.

The most plausible explanation for the fact that the melodic contour is discriminated in one pattern and not the other is that sequential organization processes operating on the basis of timbre and spatial position do not give perceptual access to a discriminable contour in the 2/2 configuration, but do in the 3/1 configuration. This conclusion is weakened somewhat by the lack of significant interaction in the across-experiment ANOVA, but the results of the planned contrasts are consistent with the stream-segregation hypothesis.

V. EXPERIMENT 4

Experiment 4 was conducted to verify that adult performance on the same stimuli in an explicit discrimination task would give similar results to those obtained with the newborns.

A. Method

In the experiments with the newborns, stimulus presentation was contingent upon high-amplitude sucking. The time interval between the last presentation of the habituation sequence and the first post-habituation sequence varied from 0.3 to 45.3 s across all three experiments ($M=11.2$ s, s.d.$=11.7$ s). The upper limit on this interval would be constrained by our low-rate rejection criterion. In order to simulate the same kind of variation for adult listeners, we therefore decided to present inter-sequence silences (ISIs) of 5, 15, and 25 s for each experimental condition. These values are longer than 37%, 70%, and 84%, respectively, of all ISIs across the habituation point in the newborn experiments.

The experiment was conducted in two 45-min sessions, one with the fast tempo/small pitch interval condition and one with the slow tempo/large interval condition. The order of presentation of the sessions was counterbalanced across two groups of six nonmusician listeners. In each session 48 trials were presented composed of two configurations (3/1, 2/2), three inter-sequence silences (5, 15, 25 s), four comparisons (rising/rising, falling/falling, rising/falling, falling/rising), and two repetitions.

The subjects heard a warning signal followed 2 s later by the initial sequence which faded in over 2 s, played 10 cycles at full level, and then faded out over 2 s. After a variable silence, the second sequence was presented in the same manner. The subject was to judge whether the order of the pitches (i.e., the pitch contour) in the two sequences was

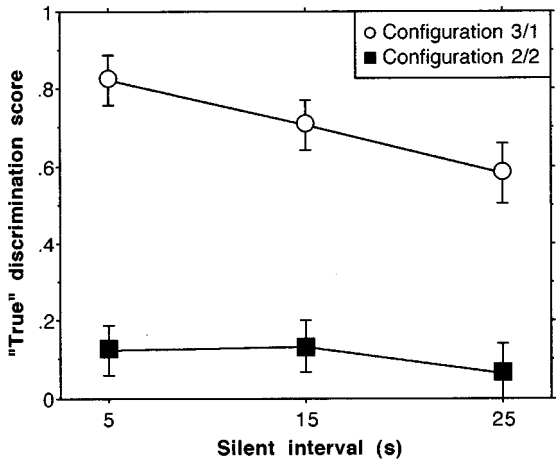


FIG. 5. Results for experiment 4. Mean ''true'' discrimination scores (hit rate minus false alarm rate) are shown as a function of duration of the silent interval separating the two sequences to be compared in a trial. (Chance performance is zero for this score.) The data for the two stimulus configurations (3/1 and 2/2), averaged across tempo/interval size conditions, are shown as separate curves. Vertical bars represent ± one standard error.

the same or different. One ''same'' and one ''different'' pair were presented for each initial sequence in each experimental condition. No feedback was given concerning the correct response.

The subjects in each group were tested collectively. They were seated in a sound-treated room in front of two loudspeakers that formed an azimuthal angle of between 60° and 90° depending on the subject's distance from the loudspeakers. The trumpet sound was always presented in the right speaker and the vibraphone in the left. Subjects marked their responses on an answer sheet.

B. Results and discussion

For each condition and subject, a ''true discrimination'' score was computed by subtracting the false-alarm rate from the hit rate across comparisons and repetitions. This gives a performance score that varies between 0 (chance performance) and 1 (perfect performance) unless listeners systematically respond incorrectly in which case the score can be negative. The mean scores across tempo/interval and presentation order are shown in Fig. 5 as a function of intersequence silence. The scores were submitted to a four-way ANOVA with repeated within-subjects factors tempo/interval(2) × configuration(2) × silence(3) and between-subjects factor presentation order of sessions (2). The only significant effects were silent interval [$F(2,20)=3.94$, $p<0.05$], indicating that performance decreases slightly overall when the delay between target and comparison sequence is long, and timbre/space configuration [$F(1,10)=74.48$, $p<0.0001$], demonstrating clearly that subjects perform well with the 3/1 configuration and very poorly with the 2/2 configuration (0.70 vs 0.11 globally), as was hypothesized at the beginning. Thus as measured in a paradigm requiring an explicit response from adult subjects, results similar to those obtained with newborns are obtained, suggesting in both cases that the sequences are organized into two streams on the basis of timbre and spatial location of the sound events.

VI. GENERAL DISCUSSION

These results show that newborns can discriminate rising and falling melodic patterns under certain conditions. They are able to make the discrimination, as measured with the non-nutritive sucking paradigm, for large pitch interval/slow tempo sequences organized in the 3/1 configuration (experiment 2), but they do not discriminate similar patterns with small interval and fast tempo conditions (experiment 1). Adult subjects easily discriminate the 3/1 patterns in both conditions using a classic 2AFC task (experiment 4). Although different paradigms were used for adults and newborns, the differences in performance indicate that melodic pattern discrimination has some limits at birth that improve with age. The available literature has not measured frequency and temporal resolution in newborns, but data from 3-month-olds suggest that the limit in our study is probably related more to poor temporal resolution that to poor frequency resolution. Our study extends work on pattern perception in older infants by Trehub *et al.* (1987), indicating that at least some of the capacities demonstrated in their study are present at birth.

Neither newborns nor adults discriminated original from retrograded 2/2 configurations (experiments 3 and 4, respectively). This result, taken together with the capacity to discriminate similar 3/1 configurations in experiment 2, suggests that the ability to discriminate melodic sequences depends on the way the events are organized into streams on the basis of timbre and/or spatial position of the sources. Work by Hartmann and Johnson (1991) on adults using an interleaved melody recognition task, indicates that it is probably the timbre difference that is primarily responsible. Our study extends work reported by Demany (1982) and Fassbender (1993) for similar kinds of sequences presented to infants of 1.5–5.5 months of age. We are inclined to interpret the data as indicating that stream organization mechanisms are present at birth. However, the statistical weakness of the critical comparisons makes it clear that this kind of study needs to be replicated with other response paradigms and perhaps performed with a paradigm in which discrimination is required to demonstrate streaming to confirm the present results in which streaming is inferred from a lack of discrimination.

This study is the first to investigate stream organization in newborns. Methodological limitations of research with newborn babies could be partly responsible for the lack of work on such questions. Most experiments address questions of the type ''Are babies able to discriminate A from B?,'' A and B being single stimulus events or categories. Questions about how newborns construct percepts and organize complex stimuli are far more difficult to address, as answers to them are often inferred from discrimination paradigms. These latter present difficulties for long sequences even in adults, as is witnessed by the relative lack of such research in this journal. In spite of the tenuousness in interpreting newborn data on such complex perceptual processing, the present data suggest, in accord with previous studies, that the most

basic processes for stream organization are operative very early in life. Babies can build auditory streams on the basis of frequency proximity, intensity similarity, and spectral similarity (Demany, 1982; Fassbender, 1993). The present study suggests that, in addition, complex timbral and spatial properties of sound sources may also be used as cues to form auditory streams from sequences of events and to differentiate the streams thus formed. These presumably unlearned processes (even if they are more affected by certain qualities of the stimuli such as pitch interval and tempo than corresponding processes in adults) would allow infants to perceptually structure their acoustic environment.

We know that infants perceive visual objects and events according to innate constraints that allow them to organize their knowledge of objects and events earlier than they could do on the basis of their own experience with these objects (Baillargeon *et al.*, 1990; Spelke *et al.*, 1992). There is almost no comparable data on the way infants acquire knowledge about the *acoustic* world. Some innate predisposition seems to be at work for perceiving speech (Pinker, 1994). But what about more general-purpose mechanisms? To what extent is the human auditory system innately predisposed to use different acoustic cues in such a way that auditory scenes could be analyzed in terms of objects or coherent sources and not as a collection of acoustic dimensions that the infant would have to learn to relate and to combine through experience? Our results are compatible with the hypothesis that some basic processes are in place at birth, even if not completely developed. More research is needed to better understand how these basic processes participate in the early development of more ''strategic'' or ''heuristic'' processes such as those that are characterized as ''schema-based'' by Bregman (1990). For example, Newman and Jusczyk (1996) recently tested 7.5-month-old infants' abilities to extract speech information (words) delivered by a voice in the presence of another, competing voice, speaking simultaneously at different intensities. Several experiments showed that the infants could recognize target words, indicating that they were able to segregate two speech streams and to selectively and continuously attend to the target speech, at least when it was more intense than the background speech.

Given the extensive research with adults on such questions, it is imperative to explore infants' primary abilities in auditory cognition. The present research will perhaps have had the merit of inciting further work in this little explored domain.

ACKNOWLEDGMENTS

This work was supported in part by a fellowship from the Fyssen Foundation to SM tenured at the Laboratoire de Sciences Cognitives et Psycholinguistique. We would like to thank Jacques Mehler for having suggested the idea of doing this research and for numerous illuminating discussions. Jacqueline Bobrow helped run subjects, and Laurent Demany and two anonymous reviewers made several helpful comments.

[1]While we conceive of the dimension of discrimination as being ''contour'' in this study, it should be pointed out that we cannot separate melodic interval pattern discrimination from discrimination of the contour of ups and downs in the melody here since the two covary, i.e., we do not present similar contours that vary in interval pattern and absolute pitch as did Trehub *et al.* (1987).

Baillargeon, R., Graber, M., DeVos, J., and Black, J. C. (**1990**). ''Why do young infants fail to search for hidden objects?,'' Cognition **36**, 255–284.

Bregman, A. S. (**1978**). ''Auditory streaming is cumulative,'' J. Exp. Psychol. **4**, 380–387.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

Bregman, A. S., and Campbell, J. (**1971**). ''Primary auditory stream segregation and perception of order in rapid sequences of tones,'' J. Exp. Psychol. **89**, 244–249.

Bregman, A. S., Liao, C., and Levitan, R. (**1990**). ''Auditory grouping based on fundamental frequency and formant peak frequency,'' Can. J. Psychol. **44**, 400–413.

Clifton, R. K. (**1992**). ''The development of spatial hearing in human infants,'' in *Developmental Psychoacoustics*, edited by L. A. Werner and E. W. Rubel (American Psychological Association, Washington, DC), pp. 135–158.

Demany, L. (**1982**). ''Auditory stream segregation in infancy,'' Inf. Behav. Dev. **5**, 261–276.

Fassbender, C. (**1993**). *Auditory Grouping and Segregation Processes in Infancy* (Kaste Verlag, Norderstedt).

Ferland, M. B., and Mendelson, M. J. (**1989**). ''Infants' categorization of melodic contour,'' Inf. Behav. Dev. **12**, 341–355.

Floccia, C., Christophe, A., and Bertoncini, J. (**1997**). ''High-amplitude sucking and newborns: The quest for underlying mechanisms,'' J. Exp. Child Psychol. **64**, 175–198.

Hartmann, W. M., and Johnson, D. (**1991**). ''Stream segregation and peripheral channeling,'' Music Percept. **9**, 155–184.

Jusczyk, P. W. (**1985**). ''The high-amplitude sucking technique as a methodological tool in speech perception research,''in *Measurement of Audition and Vision in the First Year of Life: A Methodological Overview*, edited by G. Gottlieb and N. A. Krasnegor (Ablex, Norwood, NJ), pp. 195–222.

Jusczyk, P. W. (**1997**). *The Discovery of Spoken Language* (MIT, Cambridge, MA).

Marean, G. C., and Werner, L. A. (**1991**). ''Forward masking functions of 3-month-old infants,'' J. Acoust. Soc. Am. **89**, 1914(A).

McAdams, S., and Bregman, A. S. (**1979**). ''Hearing musical streams,'' Comput. Music J. **3**, 26–43.

McAdams, S., Bertoncini, J., and Bobrow, J. (**1990**). ''Organization and discrimination of repeating sound sequences by newborn infants,'' J. Acoust. Soc. Am. Suppl. 1 **88**, S91.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. (**1995**). ''Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes,'' Psychol. Rev. **58**, 177–192.

Newman, R. S., and Jusczyk, P. W. (**1996**). ''The cocktail party effect in infants,'' Percept. Psychophys. **58**, 1145–1156.

Olsho, L. W. (**1985**). ''Infant auditory perception: Tonal masking,'' Inf. Behav. Dev. **8**, 371–384.

Olsho, L. W., Koch, E. G., and Halpin, C. F. (**1987**). ''Level and age effects in infant frequency discrimination,'' J. Acoust. Soc. Am. **82**, 454–464.

Pinker, S. (**1994**). *The Language Instinct* (Morrow, New York).

Spelke, E. S. (**1990**). ''Principles of object perception,'' Cogn. Sci. **14**, 29–56.

Spelke, E. S., Breinlinger, K., Macomber, J., and Jacobson, K. (**1992**). ''Origins of knowledge,'' Psychol. Rev. **99**, 605–632.

Spetner, N. B., and Olsho, L. W. (**1990**). ''Auditory frequency resolution in human infancy,'' Child Dev. **61**, 632–652.

Trehub, S. E., and Trainor, L. J. (**1993**). ''Listening strategies in infancy: The roots of music and language development,'' in *Thinking in Sound: The Cognitive Psychology of Human Audition*, edited by S. McAdams and E. Bigand (Oxford U.P., Oxford), pp. 278–327.

Trehub, S. E., Endman, M. W., and Thorpe, L. A. (**1990**). ''Infants' perception of timbre: Classification of complex tones by spectral structure,'' J. Exp. Child Psychol. **49**, 300–313.

Trehub, S. E., Schneider, B. A., and Henderson, J. L. (**1995**). ''Gap detection in infants, children, and adults,'' J. Acoust. Soc. Am. **98**, 2532–2541.

Trehub, S. E., Thorpe, L. A., and Morrongiello, B. A. (**1987**). ''Organization processes in infants' perception of auditory patterns,'' Child Dev. **58**, 741–749.

van Noorden, L. P. A. S. (**1977**). ''Minimum differences of level and frequency for perceptual fission of tone sequences ABAB,'' J. Acoust. Soc. Am. **61**, 1041–1045.

Werner, L. A. (**1992**). ''Interpreting developmental psychoacoustics,'' in *Developmental Psychoacoustics*, edited by L. A. Werner and E. W. Rubel (American Psychological Association, Washington, DC), pp. 47–88.

Werner, L. A., and Rubel, E. W. (Eds.) (**1992**). *Developmental Psychoacoustics* (American Psychological Association, Washington, DC).

Werner, L. A., Marean, G. C., Halpin, C. F., Spetner, N. B., and Gillenwater, J. M. (**1992**). ''Infant auditory temporal acuity: Gap detection,'' Child Dev. **63**, 260–272.

Wessel, D. L., Bristow, D., and Settel, Z. (**1987**). ''Control of phrasing and articulation in synthesis,'' in *Proceedings of the 1987 International Computer Music Conference* (Computer Music Association, San Francisco), pp. 108–116.

Detection of increments and decrements in sinusoids as a function of frequency, increment, and decrement duration and pedestal duration

Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

Robert W. Peters
Division of Speech and Hearing Sciences, Department of Medical Allied Health Professions, The School of Medicine, and Department of Psychology, The University of North Carolina at Chapel Hill, North Carolina 27599-7190

(Received 26 July 1996; accepted for publication 30 July 1997)

Thresholds for the detection of increments and decrements in level of 70 dB SPL sinusoidal signals were measured as a function signal duration (10, 20, or 200 ms), pedestal duration before the signal (10 ms, 200 ms, or pedestal on continuously) and frequency (250, 1000, or 4000 Hz). The sinusoids were presented in a low-pass filtered background noise with an overall level of 68–69 dB SPL which had two purposes: (1) to mask spectral splatter; (2) to induce an adaptation effect, which caused the continuous 4000-Hz pedestal (but not the other two pedestals) to decay to inaudibility (adaptation). We were particularly interested in determining whether the difference in noise-induced adaptation across frequency would influence the pattern of results. Seven normal-hearing subjects were used. Thresholds improved with increasing frequency and with increasing duration for both increments and decrements. However, the effect of increment/decrement duration decreased with increasing frequency; at 4000 Hz thresholds were almost the same for increment durations of 10 and 20 ms. The energy of the increments at threshold increased markedly with increasing increment duration (especially from 20 to 200 ms), suggesting a dominant role for the onsets of the increments as opposed to ongoing differences in level. Increasing the pedestal duration before the increment from 10 to 200 ms slightly improved thresholds for increment and decrement durations of 10 and 20 ms. Increment thresholds were similar for the gated and continuous pedestals at all frequencies, even though the 4000-Hz continuous pedestal decayed to inaudibility. However, thresholds for 200-ms increments were somewhat lower for continuous than for gated pedestals, and supplementary experiments found a larger gated-continuous difference for pedestals presented in quiet. Making the pedestal continuous adversely affected performance for the 10- and 20-ms decrements, but not for the 200-ms decrement. We suggest that the results for decrement detection may be affected by neural long-term adaptation, although they are not clearly related to loudness adaptation. [S0001-4966(97)04511-6]

PACS numbers: 43.66.Mk, 43.66.Fe, 43.66.Ed [JWH]

INTRODUCTION

This paper is concerned with the effects of adaptation on temporal integration in the detection of brief increments and decrements in sinusoidal stimuli (Bacon and Viemeister, 1994). Our interest in this topic was partly kindled by the observation that a high-frequency continuous sinusoid presented in background noise appears to become less loud after a few tens of seconds (Wright, 1959), and may even become completely inaudible, even though the tone is 10–15 dB above its (initial) masked threshold (Carlyon and Moore, 1986). This presumably reflects a form of long-term neural adaptation (see below for details). We were interested in the possible influence of this on the detection of increments and decrements in level. In pilot experiments we found that the task of detecting an increment in level of a high-frequency continuous sinusoid in noise resembled the task of *detecting* a tone in noise; only the increment was detectable, and not the pedestal itself. Similarly, a relatively long decrement in a continuous high-frequency pedestal was heard as a brief restoration in audibility of the pedestal at the end of the decrement (i.e., a recovery from adaptation). However, this effect was weak or absent for short decrements, and it did not occur for low pedestal frequencies, since the adaptation itself appeared weak or absent at low frequencies.

In our experiment, we explored the role of adaptation by examining how thresholds for detecting an increment or decrement in level changed as a function of increment or decrement duration and as a function of the duration of the pedestal prior to the presentation of the increment or decrement. All stimuli were presented in continuous background noise which served two purposes. First, it masked spectral splatter associated with the relatively fast onsets and offsets of the increments and decrements. Second, at high frequencies it induced the adaptation effect referred to above. We discuss

next why adaptation might be expected to influence increment and decrement detection and then we briefly review relevant previous research.

Adaptation effects in auditory neurons have been studied extensively, although mainly for tones in quiet. In response to long-duration stimulation with an abrupt onset, primary auditory neurons show an initial high discharge rate followed by a decay in rate (Smith, 1988; Palmer, 1995). This is referred to as neural adaptation and it is probably partly due to depletion of neurotransmitter in inner hair cell synaptic sites (Smith, 1988). For high-spontaneous-rate cochlear afferents, three stages have been distinguished, each characterized by an exponential decay with a specific time constant. Rapid adaptation has a time constant of a few ms. Short-term adaptation has a time constant of tens of milliseconds. Long-term adaptation has a time constant of tens of seconds (Kiang *et al.*, 1965; Smith, 1977; Westerman and Smith, 1984; Javel, 1996).

Several studies have indicated that rapid and short-term adaptation in auditory nerve fibers are subtractive rather than multiplicative in nature; when an increment in level is added to an ongoing sound (the pedestal), the increase in firing rate is independent of the duration of the pedestal (Smith, 1979). The same is true for decrements in level (Smith, 1977). However, Smith *et al.* (1985) found that while this was true for spike rates averaged over relatively long time windows (10 ms), it was not true for spike rates averaged over short time windows (1 ms). In the latter case, the change in spike rate in response to decrements decreases with increasing duration of the pedestal. It is often assumed that the central auditory system requires a constant *relative* response to achieve threshold. If this is so, then a subtractive adaptation process would enhance the detection of a brief increment with increasing delay relative to the onset of the pedestal, since the response to the increment, relative to that to the pedestal, increases with delay. In contrast, if adaptation were multiplicative, the response to the increment would be changed by the same factor as the response to the pedestal. In this case, the detectability of increments would be expected to be relatively independent of delay.

The threshold for detecting a change in level is often higher when the pedestal is gated with the increment than when the pedestal is continuous (Zwicker and Fastl, 1972; Green *et al.*, 1979; Viemeister and Bacon, 1988). Subtractive adaptation provides one possible explanation for this gated-continuous difference. However, sometimes thresholds are higher in the continuous condition when the increment duration is short, especially at low levels (Green *et al.*, 1979; Carlyon and Moore, 1986), suggesting that some other process plays a role. One possibility is ''over-integration;'' for short increments, subjects may be unable to exclude the portions of the pedestal immediately preceding and following the increment.

The effect of pedestal duration prior to presentation of a signal has been extensively studied in the context of the overshoot effect (Zwicker, 1965). It has been found that the threshold for detecting a brief signal is higher when the signal is presented close to the onset of a masker than when it is presented later in the masker. This overshoot effect is greatest for very brief high-frequency signals, presented very close to the masker onset, and it depends partly on the presence of masker energy at frequencies tuned away from the signal frequency (McFadden, 1989; Schmidt and Zwicker, 1991; Oxenham and Moore, 1995; Wright, 1997). Several researchers have speculated that the overshoot effect may be influenced by rapid or short-term adaptation (Champlin and McFadden, 1989; Bacon, 1990).

We were interested primarily in the possible effects of medium- and long-term adaptation on increment and decrement detection, rather than in the short-term adaptation that may play a role in the overshoot effect. Specifically, we hoped to clarify whether long-term adaptation is subtractive or multiplicative in nature. As in experiments on overshoot, the effect of adaptation was studied in our experiment by varying the time for which the pedestal was present prior to presentation of the increment or decrement. However, our conditions were chosen *not* to be optimal for producing the overshoot effect. Specifically, the briefest increment or decrement was longer than optimal for producing the overshoot effect, the pedestal was turned on gradually (50-ms raised-cosine envelope) and the pedestal did not contain substantial energy away from the signal frequency (apart from the background noise). Hence, we expected that any temporal effects revealed by our study would be based on different processes than those involved in producing the overshoot effect.

The effects of adaptation on increment detection have been studied previously by Scharf *et al.* (1992). They compared increment detection for two types of stimuli, a 4-kHz sinusoid and a broadband white noise. Each stimulus was presented at either 10 dB sensation level (SL) or 60 dB SL and the pedestal was either continuous or intermittent (lasting 3.1 s and presented every 5 s). Loudness adaptation was estimated by the method of successive magnitude estimation. Loudness declined substantially for the 4-kHz continuous pedestal at 10 dB SL, but declined less for the intermittent pedestal. Loudness adaptation was small for all stimuli at 60 dB SL, but was slightly larger for the 4-kHz pedestal than for the noise pedestal. For all of the stimuli, increment detection was better for the continuous pedestal than for the intermittent pedestal, suggesting that adaptation may lead to improved performance.

In further experiments, Scharf *et al.* (1992) studied how increment detection was affected by the delay of the increment relative to the onset of a continuous pedestal. For a 4-kHz pedestal, performance improved with increasing delay up to about 30 s, with a pedestal at either 10 or 50 dB SL. For a broadband noise, there was little effect of delay at either 10 or 50 dB SL. This general pattern of results is consistent with the hypothesis that adaptation may lead to improved increment detection; the improvement with increasing delay was greater for the 4-kHz tone, which showed reasonably strong adaptation, than for the noise, which did not. However, Scharf *et al.* pointed out that the effects of delay time for the 4-kHz pedestal were greater at 50 dB SL than at 10 dB SL, whereas adaptation is greater for the latter. Also, increment detection did not improve for delays beyond about 30 s, while loudness adaptation extends over at least 2–3 min. Scharf *et al.* concluded that ''intensity discrimina-

tion and loudness adaptation are, for the most part, unrelated.''

The effect of adaptation on the detection of increments has also been studied by Bacon and Viemeister (1994). They measured intensity discrimination for gated 16-kHz tone bursts and for the detection of increments in level in a continuous 16-kHz pedestal. The continuous pedestal was reported to decay to inaudibility over a period of about 30 s, and testing started after this had occurred.The thresholds were lower for the continuous than for the gated pedestal at moderate sensation levels (SLs), but not at very low SLs. In a previous study (Viemeister and Bacon, 1988) using the same two subjects they had found similar results at 1 kHz, although in this case the continuous pedestal did not become inaudible. Bacon and Viemeister suggested that their results were more consistent with a multiplicative model of (long-term) adaptation than with a subtractive model.

Substantial loudness adaptation for continuous tones in quiet occurs only for levels close to threshold, except for high-frequency tones, when it also occurs at moderate sensation levels (Scharf, 1983; Miskiewicz *et al.*, 1993; Hellman *et al.*, 1997). However, pilot experiments in our laboratory indicated that a continuous pedestal presented in background noise of moderate level would typically decay to inaudibility even at relatively high overall sound-pressure levels. The effect was very distinct for frequencies of 4 kHz and above.[1] However, it did not occur for medium to low frequencies (1 kHz and below). The adaptation to higher frequencies occurred over a wide range of overall levels, provided the pedestal was not more than about 20 dB above its (initial) masked threshold. Hence, we thought that the use of pedestals in a moderate level of background noise would allow us to study the effects of adaptation on increment and decrement detection under more typical stimulus conditions, using moderate levels and frequencies well within the audible range. By comparing performance for a frequency where strong loudness adaptation occurred (4000 Hz) and for frequencies where it did not (250 and 1000 Hz), we hoped to clarify the influence of loudness adaptation on the detection of increments and decrements.

A second goal of the experiment was to provide a further test of the ''change-detection'' hypothesis proposed by Macmillan (1971, 1973). He suggested that the auditory system contains a change detector that is insensitive to the direction of a change (i.e., whether it is an increment or a decrement), and whose sensitivity does not change with signal duration. His evidence for this came partly from studies showing that thresholds for detecting an increment or decrement in level of a sinusoid changed only slightly with increment or decrement duration. Some of Macmillan's work was criticized on the grounds that subjects may have detected spectral splatter associated with the abrupt onsets of the increments and decrements (Leshowitz and Wightman, 1972). Although Macmillan (1973) subsequently showed that not all of his results could be accounted for by splatter detection, nevertheless, it may have influenced his results obtained with sinusoidal pedestals. Our experiment was intended to provide further evidence relevant to the change-detection hypothesis by determining how thresholds for detecting increments and decrements in sinusoids change with duration, under conditions where the detection of spectral splatter was very unlikely.

I. METHOD

A. Procedure

All thresholds were measured using an adaptive three-alternative, forced-choice procedure. Each trial consisted of three observation intervals, marked by lights, and feedback was provided by lights on the response box. The ''signal,'' a decrement or increment in the level of a sinusoid, occurred randomly in one of the three intervals, and the task was to indicate the interval that contained the signal. After three consecutive correct responses, the task was made one step harder, by decreasing the magnitude of the decrement or increment; after each incorrect response the task was made one step easier. This procedure tracks the point on the psychometric function corresponding to 79.4% correct. The transition from increasing to decreasing difficulty, and vice versa, defined a turnaround. Testing continued until eight turnarounds occurred. The programmable attenuator controlling the level of the increment or decrement (see below) was changed in steps of 4 dB up to the first two reversals, and in steps of 2 dB thereafter. The mean attenuator setting at the last six reversals was calculated. This value was then used to calculate the change in level at threshold, ΔL. ΔL is equal to the difference in level of the pedestal alone and the pedestal plus increment or decrement.[2] Three estimates were obtained for each condition unless the first two estimates of ΔL were with 1 dB of each other (0.5 dB when the value of ΔL was less than 2 dB), in which case only two estimates were obtained. The final estimate was taken as the geometric mean of the two or three estimates (as the variability of the estimates of ΔL was roughly proportional to ΔL). Subjects were tested individually in a sound attenuating chamber.

B. Stimuli

The sinusoidal pedestal, with frequency 250, 1000, or 4000 Hz, was generated by a Wavetek 182 function generator. In two conditions it was gated on in three bursts each with 50-ms rise/fall time (Wilsonics BSIT, set to cosine squared, giving a raised-cosine envelope). An increment or decrement in level was presented in one of the three bursts, selected at random on each trial. The steady state duration before the increment or decrement (hereafter called the ''fringe'' duration) was either 10 or 200 ms, and the steady-state duration after the increment or decrement was 100 ms. The interval between bursts of the pedestal was 200 ms. Thus the interval between the end of one signal and the start of the next was either 410 ms or 600 ms. In a third condition, the pedestal was presented continuously. The interval between the three times when the signal might occur was 590 ms. The level of the pedestal was 70 dB SPL, specified in terms of eardrum sound pressure as measured by a probe microphone placed close to the eardrum.

A second Wilsonics gate was used to produce a burst of a sinusoid derived from the same function generator. The burst was amplified by 10 dB, passed through a Wilsonics

programmable attenuator (PATT) and added, via an active adder, either in phase to the longer burst, to produce an increment in level, or in antiphase, to produce a decrement. The attenuator was used to control the magnitude of the increment or decrement. The step size described in the procedure section refers to the setting of this attenuator. For decrements, the level of the brief burst was not allowed to exceed the level of the longer burst. If the adaptive procedure called for a greater level of the brief burst, the program provided a warning, and the run was terminated. For both increments and decrements, the brief burst had raised-cosine rise/fall times of 5 ms. The gating was done without regard to the phase of the sinusoid. The steady-state durations of the increments and decrement were 10, 20, or 200 ms. The 5-ms rise/fall time was chosen to allow brief increments or decrements while reducing the detectability of spectral splatter.

As a further means to prevent the detection of splatter, and also to induce the adaptation effect described in the introduction, the sinusoid was presented in a background of continuous white noise (Grason-Stadler, 901B) low-pass filtered at 750, 2000, and 8000 Hz (Khron-Hite 3550, 48 dB/oct slope) for the pedestal frequencies of 250, 1000, and 4000 Hz, respectively. The level of the noise was chosen to be sufficient to mask splatter while leaving the sinusoid well above masked threshold. The spectrum level of the noise was 30.5, 34.6, and 40 dB below the level of the pedestal for the frequencies 250, 1000, and 4000 Hz, respectively. The overall noise level at the output of the auditory filters was about 13.4 dB below the pedestal level for each center frequency, assuming that the auditory filter equivalent rectangular bandwidths have the values suggested by Glasberg and Moore (1990), namely 51.7, 133, and 457 Hz at 250, 1000, and 4000 Hz, respectively. At the signal-to-noise ratios used, subjects reported that no clicks or thuds were audible, and that the brief decrements or increments were heard as brief fluctuations in level of the pedestal (or brief restorations of audibility of the pedestal in cases where the pedestal decayed to inaudibility). They also reported that the 4-kHz continuous pedestal decayed to inaudibility after a period ranging from about ten seconds up to one minute. The decay to inaudibility usually occurred before the step size in the adaptive procedure decreased to 2 dB. Thus for the part of the run upon which thresholds were based, the continuous 4-kHz pedestal was inaudible.

The signal and noise were combined in an active adder and passed through a final manual attenuator to one earpiece of a Sennheiser HD424 headset, chosen for its relatively smooth frequency response. Stimulus timing and levels were controlled via an IBM PC-compatible computer that also recorded subjects' responses and provided feedback.

C. Subjects

Seven subjects were used, all with absolute thresholds below 20 dB HL at all frequencies tested. Two were authors BM and RP, aged 49 and 73, respectively. The mean age for the remaining five subjects was 23.2 years with a standard deviation of 2.2 years and a range of 5 years. Data for BM and RP were not different from those for the other subjects. All subjects had considerable prior experience in similar psychoacoustical listening tasks and were give several practice runs to familiarize them with the task. No significant practice effects were noted during the course of the experiment.

II. RESULTS

The pattern of results was generally similar across subjects, so in what follows only mean data will be presented. The means and standard deviations across subjects in all conditions are given in Table I.

A. Temporal integration for increments

Figure 1 shows the thresholds for increment detection, expressed as the change in level at threshold, ΔL, plotted as a function of increment duration with fringe duration as parameter; each panel shows results for one frequency. For each frequency, and for each fringe duration, performance improved with increasing increment duration. However, the improvement decreased with increasing frequency and was small at 4000 Hz. Indeed, at 4000 Hz, thresholds were almost identical for increment durations of 10 and 20 ms, for all fringe durations. Performance generally improved with increasing frequency.

A within-subjects analysis of variance (ANOVA) was conducted with factors frequency, increment duration, and fringe duration. All of the main factors were highly significant ($p<0.002$). The main effect of fringe duration reflects the fact that there was a small improvement in performance when the duration of the fringe was increased from 10 to 200 ms. This effect occurred for the 10- and 20-ms increments, but not for the 200-ms increment, and this was reflected in a significant interaction between increment duration and fringe duration [$F(4,24)=21.0$, $p<0.001$].

The interaction of frequency and increment duration was significant [$F(4,24)=19.8$, $p<0.001$], reflecting the fact that the improvement in threshold with increasing duration decreased with increasing frequency. The interaction of frequency and fringe duration was also significant [$F(4,24)=4.4$, $p=0.008$], reflecting the fact that the improvement in performance when the fringe duration before the increment was increased from 10 to 200 ms was less at 4000 Hz than at the two lower frequencies.

Although the interaction of increment duration and fringe duration was significant, this interaction accounted for only 2% of the variance in the data. To a first approximation, the changes in threshold with increment duration were independent of the fringe duration. Therefore, to examine the overall effects of temporal integration, it seems reasonable to collapse the data across fringe conditions. The mean values of ΔL for each increment duration are given in Table II.

The appropriate units for expressing the threshold values are not entirely clear. Leshowitz and Raab (1966) proposed that an appropriate measure was the increase in energy produced by the in-phase addition of the increment to the pedestal. They found that the detectability of increments in a continuous 1-kHz pedestal was roughly independent of increment duration when the energy in the increment was held constant. To calculate the increment in energy for our stimuli, the values of ΔL were first converted to values of $\Delta I/I$, where I is the intensity of the pedestal and ΔI is the

TABLE I. Means and standard deviations across subjects for the thresholds in each condition. Thresholds are expressed as the change in level at threshold, ΔL.

Increment/ Decrement duration		Pedestal steady state before increment or decrement					
		10 ms		200 ms		Continuous	
		Increment	Decrement	Increment	Decrement	Increment	Decrement
Frequency 250 Hz							
10 ms	mean	9.20	−15.34	8.07	−11.75	7.71	−18.15
	s.d.	2.10	2.90	1.48	1.46	1.66	2.22
20 ms	mean	7.76	−10.68	6.00	−9.24	6.38	−19.17
	s.d.	1.81	2.85	1.05	1.40	1.00	5.01
200 ms	mean	3.48	−4.81	3.51	−4.32	2.85	−6.05
	s.d.	0.80	1.59	1.37	1.17	0.86	1.45
Frequency 1000 Hz							
10 ms	mean	6.33	−8.46	5.04	−6.86	5.65	−14.31
	s.d.	1.26	1.26	0.65	0.80	1.26	4.10
20 ms	mean	5.50	−5.61	4.21	−4.70	4.80	−9.18
	s.d.	0.78	1.59	0.99	0.79	1.02	2.17
200 ms	mean	3.00	−4.22	2.96	−3.93	2.23	−3.31
	s.d.	1.09	0.47	0.74	1.08	0.83	1.15
Frequency 4000 Hz							
10 ms	mean	4.80	−5.87	3.84	−5.95	3.82	−12.63
	s.d.	0.94	1.26	0.86	1.67	0.58	3.39
20 ms	mean	4.21	−4.34	3.59	−3.94	3.06	−10.49
	s.d.	0.83	1.27	0.84	0.85	0.75	4.51
200 ms	mean	2.50	−3.62	2.66	−3.76	1.33	−2.78
	s.d.	0.82	1.04	0.62	1.51	0.48	1.09

change in intensity produced by addition of the increment to the pedestal. The values of $\Delta I/I$ were then multiplied by the equivalent rectangular duration of the increment (13.75, 23.75, and 203.75 ms, for increments with steady state durations of 10, 20, and 200 ms, respectively). The results of these calculations are also shown in Table II.

In contrast to the suggestion of Leshowitz and Raab, our thresholds are not constant when expressed in terms of the increment in energy. The energy at threshold increases with increasing increment duration, indicating imperfect energy integration. The results of Macmillan (1973) for a continuous 1-kHz pedestal also showed imperfect energy integration. It seems plausible that, under the conditions of our experiment, the onsets of the stimuli played an important role, and that energy following the onsets had only a minor influence. The origin of the discrepancy with the report of Leshowitz and Raab is not clear. At least some of their stimulus conditions were similar to ours, employing a continuous pedestal and a broadband background noise. They reported two experiments. In one, the increment energy was held constant as increment duration was varied (although details of how they did this are not given). Performance was roughly independent of increment duration. In another experiment, where the procedure was similar to ours, increment thresholds were measured as a function of increment duration. The pattern of their results in that experiment was similar to ours, although their thresholds were slightly lower, as their procedure tracked a lower value of the detectability index, d'. Analysis of the results from that experiment (their Fig. 2) indicates that the increment energy at threshold increased with increasing increment duration, just as in our results.

The ratio of increment energy for the 200-ms increment and the 10-ms increment increased with increasing frequency, indicating less effective temporal integration with increasing frequency. We conclude that, for continuous pedestals in broadband background noise, detection of increments is determined more by stimulus onsets than by ongoing stimulus energy. Onsets become relatively more important as frequency increases.

B. Temporal integration for decrements

Figure 2 shows the thresholds for decrement detection, expressed as the change in level at threshold, ΔL, plotted as a function of decrement duration with fringe duration as parameter; each panel shows results for one frequency. In some conditions, particularly for the short decrements at 250 Hz, the task was very difficult; adaptive runs were often aborted since the adaptive procedure called for a decrement magnitude greater than the largest possible value. This difficulty probably occurred for two reasons; the limited temporal resolution of the auditory system, and the fact that the effective maximum decrement value was limited to about -13 dB by the background noise. Where thresholds could not be obtained, the values of ΔL were assigned a nominal value of -20 dB. However, when examining the figures, threshold values less than about -13 dB should be treated with caution.

For the gated pedestals, thresholds clearly improved with increasing decrement duration at 250 Hz, but the improvement decreased with increasing frequency. For the continuous pedestal, there was little change in threshold when

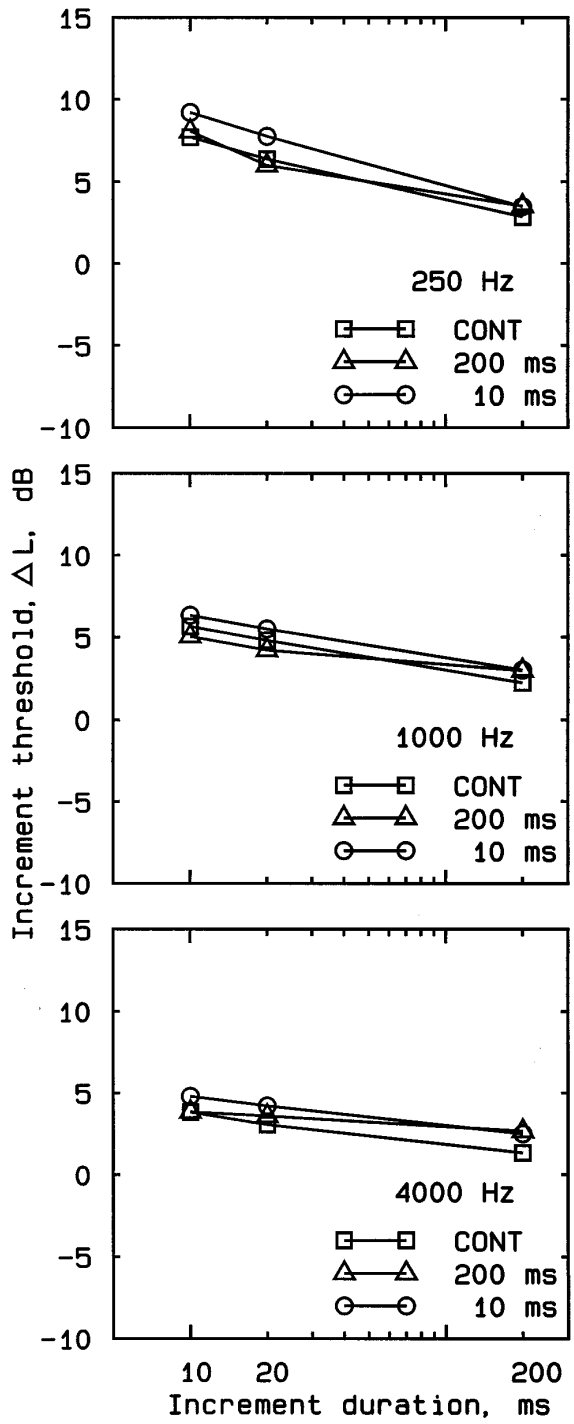


FIG. 1. Thresholds for increment detection expressed as the change in level at threshold, ΔL, plotted as a function of increment duration with fringe duration as parameter. Each panel shows results for one frequency.

the decrement duration was increased from 10 to 20 ms, but a dramatic improvement when the duration was increased to 200 ms. This pattern was observed at all three frequencies.

A within-subjects ANOVA was conducted with factors frequency, decrement duration, and fringe duration. All of the main factors were highly significant ($p<0.001$). The interaction of frequency and fringe duration was significant [$F(4,24)=22.0$, $p<0.001$], reflecting the fact that thresholds changed more with duration at 250 Hz than at higher frequencies. The interaction of frequency and fringe duration

TABLE II. Results for increment detection averaged across the different fringe durations. The first column shows the frequency. The second column shows the increment duration. The third column shows the change in level, ΔL, produced by the increment at threshold. The fourth column shows the change in intensity produced by the increment at threshold, relative to the intensity of the pedestal, i.e., $\Delta I/I$ (linear units). The fifth column shows the energy increment caused by the increment in level, expressed relative to the energy of a 1-s increment with $\Delta I/I=1$.

Frequency Hz	Increment duration, ms	ΔL dB	$\Delta I/I$ Linear	Energy in increment
250	10	8.3	5.8	0.080
	20	6.7	3.7	0.088
	200	3.3	1.1	0.230
1000	10	5.7	2.7	0.037
	20	4.8	2.1	0.049
	200	2.7	0.9	0.178
4000	10	4.2	1.6	0.022
	20	3.6	1.3	0.031
	200	2.2	0.6	0.130

was significant [$F(4,24)=3.4$, $p=0.024$]. This reflects the finding that, for the gated pedestals, performance improved when the fringe duration before the decrement was increased from 10 to 200 ms at 250 Hz and 1000 Hz, but not at 4000 Hz. The interaction of decrement duration and fringe duration was significant [$F(4,24)=22.9$, $p<0.001$], reflecting the fact that performance with the continuous pedestal was worse than with the gated pedestals for the 10- and 20-ms decrements, but not for the 200-ms decrement.

The substantial effect of decrement duration with the continuous pedestal probably reflects adaptation processes, and we discuss it in the next section. The effects of decrement duration for the gated pedestals probably reflect more familiar temporal integration processes. To examine these effects, the data were averaged for the two conditions with gated pedestals. The results are given in Table III. The table also shows the values of $\Delta I/I$ and the decrement in energy at threshold, calculated in a similar way as for Table II.

The decrement in energy at threshold increases with increasing decrement duration, indicating imperfect energy integration. The pattern of results is generally similar to that for increments, suggesting again that the transient changes in the stimuli were more important for detection than the ongoing portions of the decrements. Unlike the results for increments, the ratio of decrement energy for the 200-ms decrement and the 10-ms decrement did not increase markedly with increasing frequency. However, this may reflect a ceiling effect, since the value of $\Delta I/I$ for the 10-ms decrement at 250 Hz approached the maximum possible value of 1; the threshold in this case was indeterminate, as the noise floor limited the depth of the decrement.

In summary, as was the case for increment detection, the results indicate that thresholds are not determined by the energy in the decrement. Rather, thresholds appear to be determined more by transient changes in the stimuli, the ongoing portions of the decrements playing only a small role.

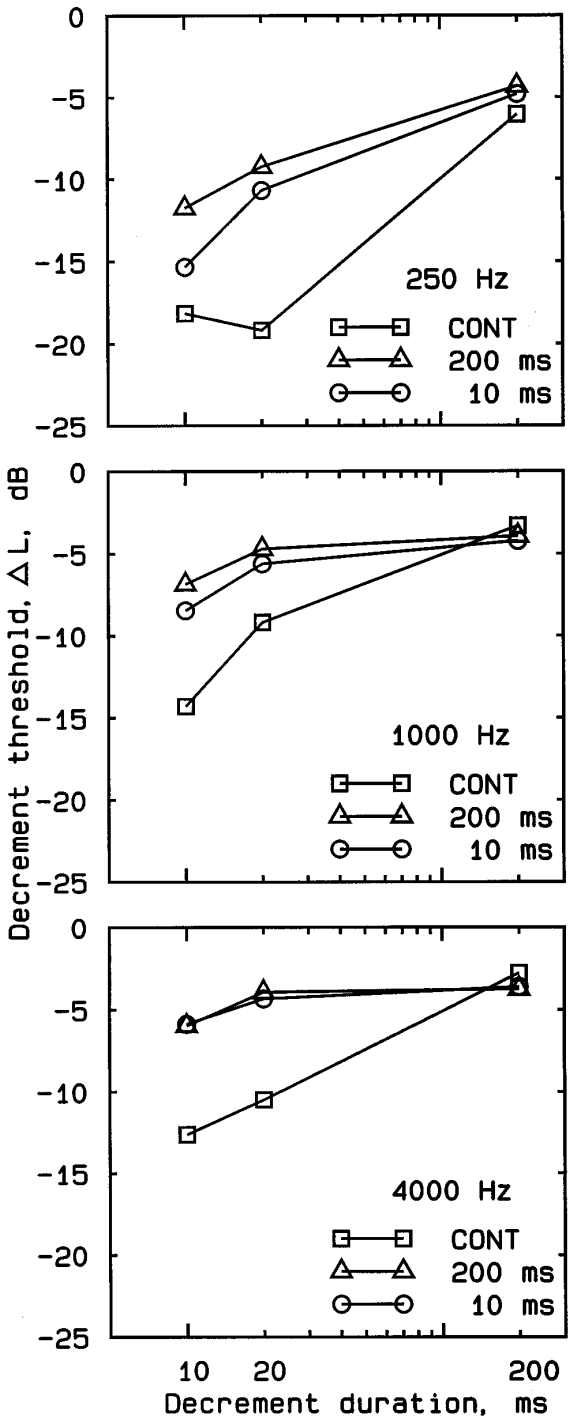


FIG. 2. As Fig. 1, but showing the results for decrement detection.

C. Effects of fringe duration

The effects of fringe duration are illustrated in Figs. 3 and 4, which show the same data as Figs. 1 and 2, but plotted as a function of fringe duration with increment or decrement duration as parameter. For the gated pedestals, thresholds tended to improve slightly when the fringe duration was increased from 10 ms to 200 ms, although this did not occur for the 200-ms decrements. This effect resembles the overshoot effect described earlier, but it is probably different in origin from the overshoot effect, since, as noted earlier, our stimulus conditions were not appropriate for producing overshoot. The effect was slightly larger at 250 Hz than at 4000 Hz, again suggesting that it is different from the overshoot effect.

TABLE III. As Table II, but showing results for decrements. Results were averaged only for gated pedestals.

Frequency Hz	Decrement duration, ms	ΔL dB	$\Delta I/I$ Linear	Energy in decrement
250	10	−13.5	0.95	0.013
	20	−10.0	0.90	0.021
	200	−4.6	0.65	0.132
1000	10	−7.7	0.83	0.011
	20	−5.2	0.69	0.017
	200	−4.1	0.61	0.125
4000	10	−5.9	0.74	0.010
	20	−4.1	0.61	0.015
	200	−3.7	0.57	0.117

The decrease in threshold with increasing fringe duration may reflect medium-term subtractive adaptation, as described in the Introduction. If the response to the pedestal declines over time, but the response to the increment or decrement stays roughly constant, then the ratio of increment/decrement response to pedestal response would increase with increasing fringe duration, giving a decrease in threshold, as we observed. The observed effect is, however, relatively small.

Making the pedestal continuous had little overall effect on detection of increments (see Fig. 3). This was the case even though the 4000-Hz continuous pedestal declined to inaudibility during the early part of a run. However, for the 200-ms increment duration, performance did improve slightly when the pedestal was made continuous, especially at 4000 Hz (see Table I). The results for the longer increment durations are similar in form to those of Scharf *et al.* (1992) (who used either 300-ms or 50-ms increments) and of Bacon and Viemeister (1994) (who used 200-ms increments) for pedestals presented in quiet; as described earlier they obtained better performance for continuous than for gated pedestals. However, larger effects are apparent in their data than in our data. We will return to this point later.

Making the pedestal continuous had substantial effects on decrement detection (see Fig. 4). It resulted in markedly increased thresholds for the two shortest decrement durations, but had little effect for the longest decrement duration. At 4000 Hz, thresholds actually improved slightly for the 200-ms decrement when the pedestal was made continuous. This occurred for six out of the seven subjects.

The effects at 4000 Hz are consistent with the subjective impression of the subjects. When the pedestal was continuous, it appeared to fade and become inaudible. When the decrement duration was long, the presence of the decrement was indicated by a brief restoration of audibility at the end of the decrement (a release from adaptation). This made the task easy. The 10- and 20-ms decrements seemed to be too brief to restore audibility, and this made the task very difficult. However, the pattern of results at 1000 and 250 Hz was similar to that at 4000 Hz, even though the continuous ped-

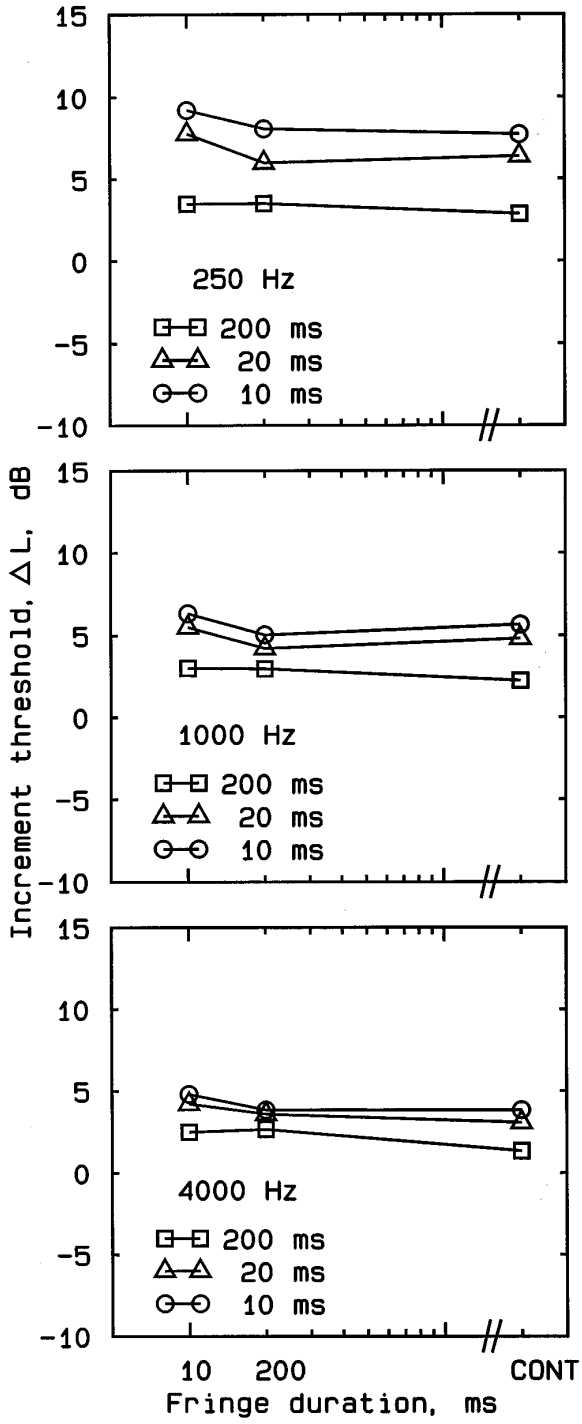


FIG. 3. Thresholds for increment detection plotted as a function of fringe duration, with increment duration as parameter.

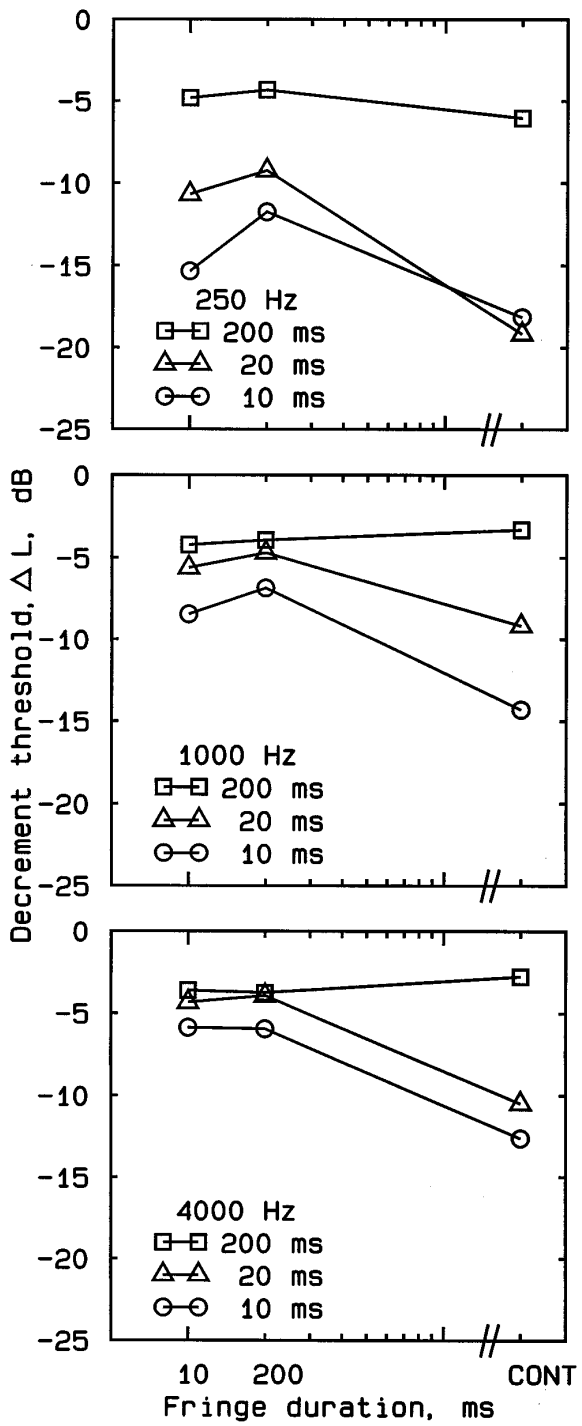


FIG. 4. As Fig. 3, but showing results for decrement detection.

estals did not decay to inaudibility at the two lower frequencies.

III. DISCUSSION

A. Effects of adaptation

Although our results show some minor differences between the results for the 4000-Hz pedestal and the results for the lower pedestal frequencies, the overall pattern of the results was similar across pedestal frequencies. Thus there appears to be little relationship between the strong loudness adaptation that occurred at 4000 Hz and performance in increment and decrement detection. This finding is consistent with the conclusion of Scharf *et al.* (1992) and of Bacon and Viemeister (1994) that increment detection is not clearly related to loudness adaptation. Neural long-term adaptation presumably did occur for the continuous pedestal, but this did not have a substantial effect on increment detection. This is consistent with the suggestion of Bacon and Viemeister (1994) that long-term adaptation is multiplicative rather than subtractive in nature.

We noted earlier that Scharf *et al.* (1992) and Bacon and

TABLE IV. Percent correct detection for increments in gated and continuous pedestals for four subjects. Each percentage is based on 100 trials for a fixed value of ΔL, which is given in the table. Subjects were tested either in quiet or in the presence of continuous background noise (see text for details).

		Quiet					Noise				
			Gated		Continuous			Gated		Continuous	
Frequency		ΔL	Percent	d'	Percent	d'	ΔL	Percent	d'	Percent	d'
250 Hz											
Subject	MD	1.0	47	0.46	86	1.97	3.0	71	1.28	85	1.91
	CA	1.2	57	0.78	96	2.85	2.4	61	0.92	77	1.52
	WH	1.0	47	0.46	86	1.97	3.0	63	0.99	83	1.80
	RP	1.5	60	0.89	88	2.09	3.0	44	0.36	79	1.61
1000 Hz											
Subject	MD	1.0	61	0.92	94	2.59	2.0	52	0.62	74	1.39
	CA	0.8	64	1.02	80	1.65	1.6	42	0.30	72	1.31
	WH	1.0	51	0.59	96	2.85	2.0	51	0.62	74	1.39
	RP	1.0	61	0.92	95	2.71	2.0	46	0.43	91	2.31
4000 Hz											
Subject	MD	0.6	44	0.36	82	1.75	1.2	49	0.52	65	1.06
	CA	0.6	34	0.02	100	∞	1.2	42	0.30	92	2.39
	WH	0.9	46	0.43	99	3.62	1.8	47	0.46	87	2.03
	RP	0.9	52	0.62	97	3.02	1.8	66	1.09	96	2.85

Viemeister (1994) found somewhat larger gated-continuous differences than we did for a 200-ms increment. The difference may have occurred because of our use of a continuous background noise in all conditions. To check on this, we repeated some of our increment-detection conditions both with and without background noise. The increment duration was 200 ms, and the pedestal was either continuous or gated on 200 ms before the increment (and off 100 ms after the increment). The pedestal level was 70 dB SPL, and background noise was either absent or present at the same levels as used in our main experiment. Measurements were obtained for four subjects (all of whom had taken part in the main experiment) for all three frequencies.

Rather than using an adaptive procedure, we used fixed values of ΔL for each subject and each frequency (as was done by Scharf *et al.*, 1992), selected on the basis of pilot experiments so that performance would be above chance for the gated pedestal but below 100% for the continuous pedestal. The values of ΔL generally had to be about twice as large when noise was present as when it was absent; the deleterious effect of the noise on performance can plausibly be explained in terms of the random fluctuations in level introduced by the background noise, and by the fact that the noise prevents the use of information from regions of the excitation pattern of the signal that are remote from the signal frequency (Florentine and Buus, 1981; Zwicker and Fastl, 1990). The values of ΔL are given in Table IV, which also shows the percent correct obtained by each subject in each condition, and the corresponding value of d'.

In the presence of noise, performance was consistently better for the continuous pedestal than for the gated pedestal. The effect was somewhat larger than expected from our main experiment, which used an adaptive procedure.[3] In the corresponding conditions of that experiment, the mean values of ΔL at 250, 1000, and 4000 Hz were 3.51, 2.96, and 2.66 dB for the gated pedestal and 2.85, 2.23, and 1.33 dB for the continuous pedestal. In the absence of noise, the gated-continuous differences were generally markedly larger than in the presence of noise. For example, for subject WH at 1000 Hz, d' values for the gated and continuous pedestals, respectively, were 0.59 and 2.85 in quiet, and 0.62 and 1.39 in noise. Similarly, for subject CA at 4000 Hz, d' values for the gated and continuous pedestals, were 0.02 and ∞ (100% correct) in quiet, and 0.30 and 2.39 in noise. The ratio of d' values for the continuous and gated pedestals, averaged across subjects and frequencies (but excluding the results for CA at 4000 Hz, where d' was infinite for the gated pedestal in quiet), was 4.1 in quiet and 2.7 in noise. It seems clear that background noise leads to a reduction in the gated-continuous difference. A possible reason is that noise may induce neural adaptation at the signal frequency even for the gated pedestal, so that the ''benefit'' of neural adaptation is present for both the gated and continuous pedestals. However, both in quiet and in noise, there was only a modest effect of center frequency on the gated-continuous difference. Once again, we are led to the conclusion that the marked loudness adaptation that occurred for the 4000-Hz pedestal in noise did not substantially affect the pattern of results.

It may be the case that loudness adaptation is related only indirectly to neural adaptation. Neural adaptation is very marked during the first few hundred ms following the onset of a stimulus; as described in the introduction, rapid and short-term adaptation can be substantial. However, loudness usually increases over the first 100 ms or so following the onset of a sound. It does not generally decline substantially during the first few hundred ms. Rather, loudness adaptation develops over several tens of seconds. This discrepancy suggests that perceived loudness is not strongly affected by short-term changes in neural firing rate in response to a sustained stimulus. Possibly, the auditory system

has some way of compensating for these changes, so as to arrive at a stable percept.

Our results for decrement detection showed large effects of the duration of the pedestal. Decrement detection was much worse for the continuous pedestal than for the gated pedestals when the decrement duration was 10 or 20 ms, but not when it was 200 ms. This finding could be explained in terms of long-term neural adaptation. The long-term firing rate in response to a continuous sinusoid in noise may decline over time to a magnitude similar to that for the noise alone. This may make the detection of brief decrements very difficult, since the decrements do not result in a marked decrease in firing rate. For the 200-ms decrement, the decrease in level during the decrement may have resulted in a release from adaptation, causing a transient increase in firing rate at the end of the decrement, which provided an effective detection cue.

If it is the case that neural firing rates in response to a continuous pedestal in noise decline over time to a magnitude similar to that for the noise alone, then we need to explain why the continuous pedestals did not decay to inaudibility at the two lower frequencies. It is possible that long-term adaptation was occurring at all frequencies. Neural long-term adaptation for tones in quiet appears to be similar for high and medium frequencies (Kiang *et al.*, 1965; Javel, 1996), although we are aware of no data indicating whether this is so when background noise is present. However, for the two lower frequencies, the adaptation in neural firing rate may not have resulted in loss of audibility because the presence of a tonal component was signaled by phase-locking information. At 4000 Hz, phase locking tends to be weak, and that, combined with the relatively low signal-to-noise ratio, may have made phase locking an insufficient cue to maintain audibility of the continuous 4000-Hz pedestal. If this argument is correct, then the detection of the 200-ms decrement appears to have been based mainly on changes in firing rate rather than on phase-locking information.

B. Temporal integration

Thresholds for increment detection changed relatively little with increment duration, especially for the 4000-Hz pedestal. This was true both for the gated and for the continuous pedestal. The energy required for threshold increased markedly with increment duration, indicating that threshold did not correspond to a constant energy increment, as proposed by Leshowitz and Raab (1966). For the gated pedestals, thresholds for decrement detection also did not change markedly with decrement duration. This was not the case for the continuous pedestal, but in this case the results were probably affected by neural long-term adaptation, as discussed above.

These results are consistent with the ''change-detection'' hypothesis proposed by Macmillan (1973). As described in the Introduction, he suggested that the auditory system contains a change detector that is insensitive to the direction of a change (i.e., whether it is an increment or a decrement), and whose sensitivity does not change with signal duration. This change detector was assumed to act in addition to a time-integrating detector that is sensitive to the direction of a change. His evidence for this hypothesis came from studies comparing the ability to detect a change in a stimulus (either an increment in level, or, in separate blocks of trials, a decrement in level) with the ability to discriminate an increment from a decrement. He found that detection performance improved less with increasing duration than did recognition performance. Hafter *et al.* (1996) have also presented data supporting the idea of a change detector. They used a task in which either an increment or decrement in a sinusoidal pedestal was presented at random on each trial. They found that subjects were better at detecting the change than they were at deciding whether it was an increment or decrement. However, this pattern was reversed when the increment or decrement was preceded and followed by changes in level. The presence of a transient change in level just prior to and following the increment or decrement seemed to disrupt the operation of the change detector.

Some of the results of Carlyon and Moore (1986) also support the idea of a change detector. They showed that the detection of a brief increment in level of a 6500-Hz continuous sinusoid was impaired by masking the onset and offset of the increment with brief bursts of bandpass noise. The impairment occurred for all levels, but was greater for pedestal levels of 55 dB and 80 dB SPL than for a level of 35 dB SPL. However, when a continuous bandstop noise was added to the continuous sinusoidal pedestal, the effect of masking the onset and offset of the increment was small. Carlyon and Moore suggested that this might be related to an adaptation effect that occurred when the bandstop noise was present. Just as in the present experiment, the continuous pedestal was reported to fade to inaudibility after a few seconds. In this case, the presence of the increment was signaled by a brief restoration of audibility. Apparently the brief bursts of bandpass noise used to mask the onset and offset of the increment did not have this effect.

Overall, the evidence for a change detector appears to be rather strong. For the detection of increments and decrements in suprathreshold stimuli it appears to be the dominant mechanism, although the ongoing portions of the increments and decrements do play some role, since performance in our experiment did improve somewhat with increasing increment/decrement duration. It remains unclear why the threshold intensity of tones at absolute threshold decreases markedly with increasing duration, while the threshold for detection of increments and decrements in suprathreshold pedestals decreases only slightly with duration. The integration apparent at absolute threshold may reflect repeated sampling of the stimuli or ''multiple looks,'' as suggested by Viemeister and Wakefield (1991). Such repeated sampling may not operate for suprathreshold pedestals, since changes in stimuli are the main carriers of information.

IV. SUMMARY

Thresholds for detecting increments and decrements in sinusoidal signals in background noise were measured. Results show:

(1) For both increments and decrements, thresholds are NOT constant when expressed as the energy in the increment or decrement.
(2) Thresholds seem to be determined more by onsets than by ongoing differences in level. This is consistent with the idea of a change-detection mechanism.
(3) Changes in threshold with increment duration are greater at lower frequencies.
(4) When the pedestal was made continuous, this had little effect on increment detection, although thresholds for detecting increments did improve slightly for the longest increment duration. Decrement detection was impaired for the short decrements, but not for the longest one (200 ms).
(5) The effect of making the pedestal continuous was similar across frequency, even though the 4000-Hz pedestal faded to inaudibility, while the other pedestals did not. This suggests that increment and decrement detection are not closely related to loudness adaptation, although decrement detection may be affected by long-term neural adaptation.
(6) Overall, the data are consistent with the idea that long-term adaptation is multiplicative rather than subtractive in nature.

ACKNOWLEDGMENTS

This work was supported by the Medical Research Council (U.K.). We thank the graduate students in Speech and Hearing at Chapel Hill for serving as listeners, and Will Hula and Deborah Vickers for assistance in the preparation of figures. We thank Peter Fitzgibbons, Joseph W. Hall, Bertram Scharf, and an anonymous reviewer for helpful comments on an earlier version of this paper.

[1]To induce complete adaptation, it may be necessary that the noise spectrum extends both above and below the signal frequency. Scharf (1983) found only partial adaptation for a 4000-Hz sinusoid presented 15 dB above masked threshold in a noise high-pass filtered at 4000 Hz.

[2]There are both advantages and disadvantages to the use of ΔL as a measure. The main advantage is that discriminability, d', is roughly proportional to ΔL (as we have found in other unpublished experiments). The main disadvantage is that the variability is not constant across different values of ΔL (see Table I); standard deviations across subjects ranged from 0.48 for a ΔL of 1.33 to 2.9 for a ΔL of -15.3. However, the pattern of the results was essentially unchanged when the measure $\Delta I/I$ was used. Also, all of the significant effects obtained in ANOVAs based on the ΔL values were also present in ANOVAs based on $\Delta I/I$.

[3]For the comparable conditions of our main experiment the mean ratio of ΔL values for the continuous and gated conditions was 0.66. In other experiments, we have found that d' for increment detection is roughly proportional to ΔL. Hence, we would expect the ratio of d' values for a fixed ΔL to be about $1/0.66 = 1.5$. In fact, the ratio had a mean value of 2.7. The gated-continuous difference may be greater when fixed increment levels are used because subjects may fail to ''home in'' on the optimal detection cues when detectability is low. This results in greater differences between conditions with high detectability and conditions with low detectability. When an adaptive procedure is used, each run starts with easy conditions, giving subjects a chance to ''find'' the appropriate detection cues, and if the cues are ''lost'', then the procedure automatically presents some more detectable stimuli (Taylor and Forbes, 1983). To check on this idea, one subject (author RWP) was tested using a procedure with five fixed increment values within a run, presented in a regular cycle. The largest increment was chosen to be highly detectable, providing a ''reminder'' of what to listen for every five trials. This method of determining psychometric functions was proposed by Moore and Sek (1992). Conditions with both gated and continuous pedestals were used. When the increment was of a moderate size, the percent correct obtained using this method was consistently greater than obtained when the same increment size was used as the single value in a block of trials. This confirms that, when a fixed increment of moderate or low detectability is used, subjects may fail to ''find'' the optimal detection cues.

Bacon, S. P. (**1990**). ''Effect of masker level on overshoot,'' J. Acoust. Soc. Am. **88**, 698–702.
Bacon, S. P., and Viemeister, N. F. (**1994**). ''Intensity discrimination and increment detection at 16 kHz,'' J. Acoust. Soc. Am. **95**, 2616–2621.
Carlyon, R. P., and Moore, B. C. J. (**1986**). ''Continuous versus gated pedestals and the ''severe departure'' from Weber's Law,'' J. Acoust. Soc. Am. **79**, 453–460.
Champlin, C. A., and McFadden, D. (**1989**). ''Reductions in overshoot following intense sound exposures,'' J. Acoust. Soc. Am. **85**, 2005–2011.
Florentine, M., and Buus, S. (**1981**). ''An excitation-pattern model for intensity discrimination,'' J. Acoust. Soc. Am. **70**, 1646–1654.
Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hearing Res. **47**, 103–138.
Green, D. M., Nachmias, J., Kearney, J. K., and Jeffress, L. A. (**1979**). ''Intensity discrimination with gated and continuous sinusoids,'' J. Acoust. Soc. Am. **66**, 1051–1056.
Hafter, E. R., Bonnel, A. M., and Gallun, E. J. (**1996**). ''Detection of change without regard to its valence,'' J. Acoust. Soc. Am. **99**, 2541.
Hellman, R. P., Miskiewicz, A., and Scharf, B. (**1997**). ''Loudness adaptation and excitation patterns: Effects of frequency and level,'' J. Acoust. Soc. Am. **101**, 2176–2185.
Javel, E. (**1996**). ''Long-term adaptation in cat auditory-nerve fiber responses,'' J. Acoust. Soc. Am. **99**, 1040–1052.
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).
Leshowitz, B., and Raab, D. H. (**1966**). ''Effects of stimulus duration on the detection of sinusoids added to continuous pedestals,'' J. Acoust. Soc. Am. **41**, 489–496.
Leshowitz, B., and Wightman, F. L. (**1972**). ''On the importance of considering the signal's frequency spectrum: Some comments on Macmillan's 'Detection and recognition of increments and decrements in auditory intensity','' Percept. Psychophys. **12**, 209–210.
Macmillan, N. A. (**1971**). ''Detection and recognition of increments and decrements in auditory intensity,'' Percept. Psychophys. **10**, 233–238.
Macmillan, N. A. (**1973**). ''Detection and recognition of intensity changes in tone and noise: the detection-recognition disparity,'' Percept. Psychophys. **13**, 65–75.
McFadden, D. (**1989**). ''Spectral differences in the ability of temporal gaps to reset the mechanisms underlying overshoot,'' J. Acoust. Soc. Am. **85**, 254–261.
Miskiewicz, A., Scharf, B., Hellman, R., and Meiselman, C. (**1993**). ''Loudness adaptation at high frequencies,'' J. Acoust. Soc. Am. **94**, 1281–1286.
Moore, B. C. J., and Sek, A. (**1992**). ''Detection of combined frequency and amplitude modulation,'' J. Acoust. Soc. Am. **92**, 3119–3131.
Oxenham, A. J., and Moore, B. C. J. (**1995**). ''Overshoot and the ''severe departure'' from Weber's law,'' J. Acoust. Soc. Am. **97**, 2442–2453.
Palmer, A. R. (**1995**). ''Neural signal processing,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).
Scharf, B. (**1983**). ''Loudness adaptation,'' in *Hearing Research and Theory, Volume 2*, edited by J. V. Tobias and E. D. Schubert (Academic, New York).
Scharf, B., Canévet, G., and Ward, L. M. (**1992**). ''On the relation between intensity discrimination and adaptation,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford).
Schmidt, S., and Zwicker, E. (**1991**). ''The effect of masker spectral asymmetry on overshoot in simultaneous masking,'' J. Acoust. Soc. Am. **89**, 1324–1330.
Smith, R. L. (**1977**). ''Short-term adaptation in single auditory-nerve fibers: Some poststimulatory effects,'' J. Neurophysiol. **49**, 1098–1112.
Smith, R. L. (**1979**). ''Adaptation, saturation and physiological masking in single auditory-nerve fibers,'' J. Acoust. Soc. Am. **65**, 166–178.

Smith, R. L. (**1988**). ''Encoding of sound intensity by auditory neurons,'' in *Auditory Function: Neurobiological Basis of Hearing*, edited by G. Edelmann, W. Gall, and M. Cowan (Wiley, New York).
Smith, R. L., Brachman, M. L., and Frisina, R. D. (**1985**). ''Sensitivity of auditory-nerve fibers to changes in intensity: A dichotomy between decrements and increments,'' J. Acoust. Soc. Am. **78**, 1310–1316.
Taylor, M. M., and Forbes, S. M. (**1983**). ''PEST reduces bias in forced-choice psychophysics,'' J. Acoust. Soc. Am. **74**, 1367–1374.
Viemeister, N. F., and Bacon, S. P. (**1988**). ''Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones,'' J. Acoust. Soc. Am. **84**, 172–178.
Viemeister, N. F., and Wakefield, G. H. (**1991**). ''Temporal integration and multiple looks,'' J. Acoust. Soc. Am. **90**, 858–865.
Westerman, L. A., and Smith, R. L. (**1984**). ''Rapid and short-term adaptation in auditory nerve responses,'' Hearing Res. **15**, 249–260.
Wright, B. A. (**1997**). ''Detectability of simultaneously masked signals as a function of masker bandwidth and configuration for different signal delays,'' J. Acoust. Soc. Am. **101**, 420–429.
Wright, H. N. (**1959**). ''Auditory adaptation in noise,'' J. Acoust. Soc. Am. **31**, 1004–1012.
Zwicker, E. (**1965**). ''Temporal effects in simultaneous masking by white-noise bursts,'' J. Acoust. Soc. Am. **37**, 653–663.
Zwicker, E., and Fastl, H. (**1972**). ''On the development of the critical band,'' J. Acoust. Soc. Am. **52**, 699–702.
Zwicker, E., and Fastl, H. (**1990**). *Psychoacoustics—Facts and Models* (Springer-Verlag, Berlin).

Binaural detection with spectrally nonoverlapping signals and maskers: Evidence for masking by aural distortion products

Marcel van der Heijden and Constantine Trahiotis
Surgical Research Center, Department of Surgery (Otolaryngology) and Center for Neurological Sciences, University of Connecticut Health Center, Farmington, Connecticut 06030

Armin Kohlrausch and Steven van de Par
Institute for Perception Research (IPO), P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands

(Received 9 January 1997; accepted for publication 24 July 1997)

Thresholds were measured for diotic tonal signals in the presence of interaurally delayed bands of Gaussian noise. When the signal frequency was 525 Hz, the spectrum of the noise was either below (highest frequency, 450 Hz) or above (lowest frequency, 600 Hz) the frequency of the signal. When the signal frequency was 450 Hz, the spectrum of the noise was always above the signal frequency (lowest frequency, 600 Hz). Signals had a 250-ms duration and were temporally centered within the 300-ms long bursts of noise. The spectrum level of the noise was 60 dB. Thresholds obtained in all three conditions varied essentially sinusoidally with the interaural delay of the noise. For signals below the spectrum of the noise, the periodicities within the data were close to, but not identical with, the periodicities of the *signals*. This outcome is discussed in terms of masking produced by aural distortion products stemming from interactions within the bands of noise [cf. van der Heijden and Kohlrausch, J. Acoust. Soc. Am. **98**, 3125–3134 (1995)]. For signals above the spectrum of the noise, the periodicities in the data suggested that masking was produced by components within the band of noise. Patterns within the data are also discussed in terms of limitations concerning the magnitude of external delays that can be matched by internal delays that are incorporated in modern models of binaural processing. [S0001-4966(97)03211-6]

PACS numbers: 43.66.Pn, 43.66.Dc [JWH]

INTRODUCTION

In a recent study, van der Heijden and Kohlrausch (1995) showed that aural distortion products evoked by interactions among the components of a band of noise determined the detectability of a tonal signal whose frequency was *below* the spectral region of the noise. The level of the noise was low enough to preclude similar effects due to remote masking (e.g., Bilger and Hirsh, 1956). Van der Heijden and Kohlrausch found that the distortion products responsible for their ''downward'' masking effects had level-dependent phases that were similar to those found with odd-order distortion products such as the cubic difference tone (e.g., Goldstein, 1967). In addition, van der Heijden and Kohlrausch found that aural distortion products appeared to play no role when the frequency of the signal was above the spectral region of the noise (an ''upward'' masking condition). Their findings are consistent with well-known physiological and psychophysical observations concerning distortion products.

It occurred to us that a binaural paradigm could provide a further test of the ability of auditory distortion products to produce ''downward'' masking. The challenge was to differentiate between masking effects produced by components of a noise stimulus per se and masking effects produced by internally generated distortion products. Our strategy was to employ interaural delays of a band of noise and to use periodicities within the data to determine whether detectability was due to: (1) components within the band of noise, in this case remote from the frequency of the signal; or (2) distortion products having spectral components in the *same* frequency region as the signal. Said differently, we attempted to use interaural delays of a band of noise to separate off-frequency masking (i.e., masking due to spectral components spectrally distant to the signal frequency) from on-frequency masking (i.e., masking due to distortion products adjacent to and/or overlapping with the signal frequency).

Our expectation was that, if masking were ''on-frequency'' due to distortion products, then the pattern of thresholds, as a function of interaural delay of the noise, would reflect the frequency of the signal. That type of outcome was reported in the classic studies of Rabiner *et al.* (1966) and Langford and Jeffress (1964), who measured the detectability of tonal signals masked by Gaussian noise having components in the same spectral region as the signal. For the type of on-frequency masking which they studied, the data showed oscillations at the frequency of the signal as the masking noise was interaurally delayed.

To our knowledge, there are no data available concerning how binaural detectability varies with interaural delay when signal and masker occupy different spectral regions. We are aware of only two studies dealing with binaural processing in ''off-frequency'' conditions. McFadden *et al.* (1972) measured masking-level differences in a tone-on-tone paradigm with the signal placed spectrally above or below the sinusoidal masker and Hall *et al.* (1983) measured binaural detection in a notched-noise experiment. Neither of those two studies incorporated stimulus conditions involving interaural delays. Therefore, the data obtained in those studies may not directly bear on issues of interest here. Conse-

quently, what the patterning of thresholds is as a function of delays in the maskers in the off-frequency case is an open question, which the data obtained in this investigation will help answer.

I. METHOD

Detection thresholds were determined using a three-interval forced-choice adaptive procedure (Levitt, 1971). Each trial consisted of three 300-ms observation intervals each separated by 200 ms. The signal occurred randomly, with equal *a priori* probability in each of the three intervals. The listeners were provided with correct-answer feedback via a computer terminal. Each estimation of threshold began with the signal being easily detectable. The level of the signal was decreased after two consecutive correct responses and was increased after each incorrect response. This procedure tracked the 70.7% correct point of the psychometric function. The step size was 8 dB at the beginning of each run, was reduced to 4 dB after the second reversal, and to 2 dB after the fourth reversal. Ten more reversals were obtained using 2-dB steps. An estimate of threshold was obtained by calculating the median of the levels over the ten last reversals. Final thresholds are the average of three such estimates.

Thresholds were obtained for tonal signals in the presence of a band of noise. In different conditions, the frequency of the signal was either below or above the frequencies composing the band of noise. When the band of noise had components ranging from 600 to 1100 Hz, the frequency of the signal was either 450 or 525 Hz. When the band of noise had components ranging from 0 to 450 Hz, the frequency of the signal was 525 Hz. Thresholds were measured with several interaural time delays (ITDs) of the noise ranging from 0 to 2470 μs in steps of 130 μs. The bands of noise had a duration of 300 ms including 20-ms, cosine-squared on/off ramps and had a spectrum level equivalent to 60 dB SPL. The duration of the signals was 250 ms (including 20-ms, cosine^2, on/off ramps) and the signals were temporally centered in the band of noise. The signals were always presented diotically (So) resulting in a binaural configuration commonly referred to as NτSo.

All stimuli were digitally generated utilizing a 32-kHz sampling rate and were played out using 16-bit D/A converters. Before each block of trials that led to an estimate of threshold, a 4-s long circular buffer of Gaussian noise was calculated and independent, 300-ms long, samples of that noise were drawn randomly to produce the noise stimulus in each observation interval. The buffers of noise were constructed by first drawing independent samples from a Gaussian distribution and then applying a discrete Fourier transform to those samples which resulted in a spectrum having a spacing of 0.25 Hz between components. The spectral components outside the desired passband were set to zero and an inverse Fourier transform yielded the desired 4-s-long circular buffer of band-limited Gaussian noise. Noises constructed in this manner have a long-term spectrum with extremely steep spectral edges.

The testing of the conditions was completely randomized, except that the estimates of thresholds obtained with signal frequencies of 450 and 525 Hz in the presence of the higher-frequency noise were collected in pairs. Three of the authors (AK, SP, and MH) participated in the experiment. All listeners had normal hearing and had extensive experience in formal binaural listening tasks. Initial testing was conducted in single-walled, sound-attenuating chambers housed in laboratories at IPO, The Netherlands, using Beyer DT 880 headphones. Follow-up experiments were conducted at IPO and at the psychoacoustics laboratory at the University of Connecticut Health Center using TDH-39 earphones at both institutions.

II. RESULTS

The behavioral thresholds are plotted in Fig. 1 along with curves representing the cosine functions that best fit the respective sets of data. In Fig. 1, panels (a) and (b), respectively, contain data obtained when the frequency was 525 Hz (closed triangles) or 450 Hz (closed squares) and when the noise had components ranging from 600 to 1100 Hz. Panel (c) of Fig. 1 contains data obtained when the frequency of the signal was 525 Hz and the components of the noise ranged from 0 to 450 Hz (closed circles).

The data points represent means calculated across the three listeners and the error bars represent the standard errors of those means. Because we are principally interested in the patterning of the data as a function of ITD, individual differences in overall threshold were removed during the calculation of the standard errors in each plot. This was done by subtracting the mean of the data of each listener from the grand mean, in each signal condition, respectively.

Let us first consider data obtained with the higher-frequency band of noise [Fig. 1, panels (a) and (b)]. In both cases, the thresholds vary with ITD in a manner that appears to be approximately cosinusoidal. Note that the patterns of thresholds obtained with the two signals are different, especially for ITDs greater than 1000 μs. When the frequency of the signal was 525 Hz [panel (a)], thresholds increased more steeply for delays greater than 1000 μs than did the thresholds obtained when the signal was 450 Hz [panel (b)]. In addition, when the signal frequency was 525 Hz, a maximum in threshold occurred when the ITD was 1690 μs and performance improved as ITDs were increased to 2470 μs. In contrast, when the signal frequency was 450 Hz, thresholds increased for ITDs of up to 2080 μs or so.

We examined periodicities within the data using the simplifying assumption that the thresholds varied as a cosinusoidal function of ITD. We found the best fitting cosine function (based on a least-squares criterion) for the data shown in panels (a) and (b). Amplitude and frequency were allowed to vary, but phase was constrained to be zero. This was done because a value of zero ITD would be expected to lead to the poorest performance (i.e., a maximum in threshold) given that the signals were presented in the So configuration. That is, the zero ITD in the noise results in an NoSo condition, a binaural configuration that does not contain interaural differences.

When the frequency of the signal was 525 Hz [panel (a)], the data were best fit by a cosine function having a frequency of 540 Hz and 84% of the variance in the data was

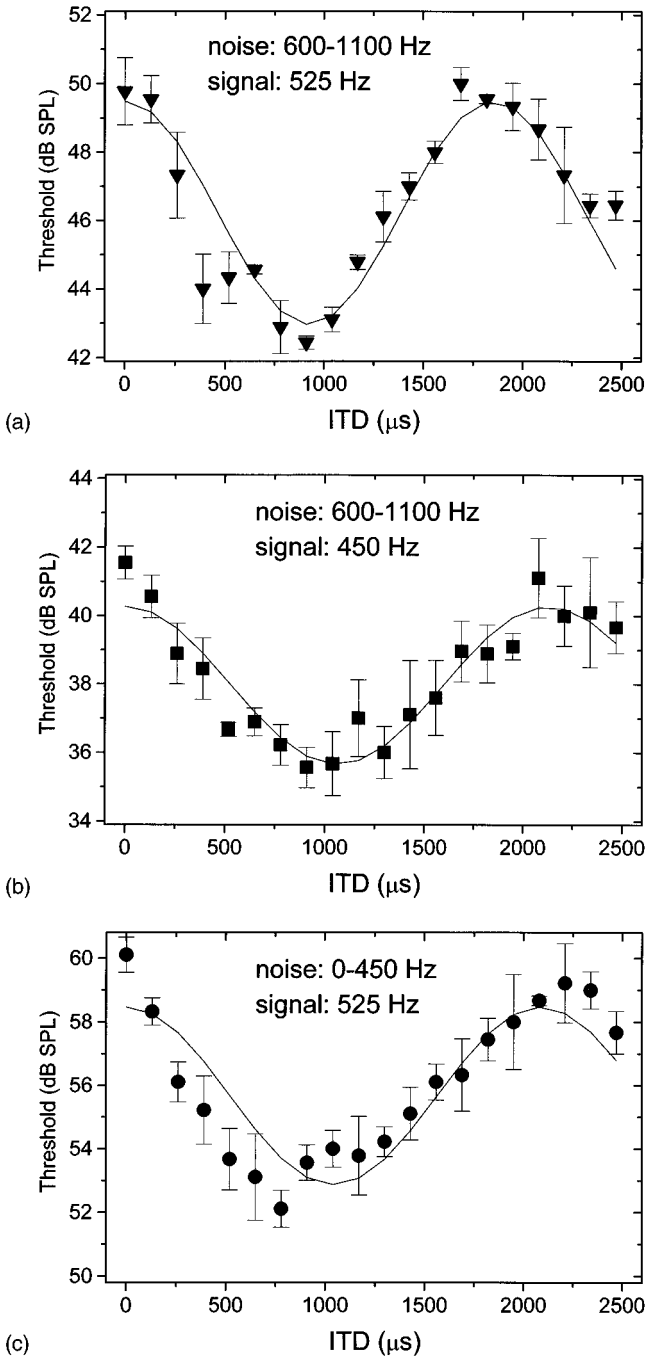


FIG. 1. Average thresholds of the three listeners as a function of ITD of the noise. Error bars indicate across-listener standard errors computed after differences in overall threshold were removed (see text). The data points in panels (a) and (b) were obtained when the spectral content of the noise ranged from 600 to 1100 Hz and when the frequency of the signal was either 525 Hz [panel (a), closed triangles] or 450 Hz [panel (b), closed squares]. The data points in panel (c) (closed circles) were obtained when the spectral content of the noise ranged from 0 to 450 Hz and the frequency of the signal was 525 Hz. Solid lines in each panel indicate cosine functions that best fit the data.

accounted for by that function. That function is represented by the solid line in panel (a). Note that this best fitting frequency of 540 Hz is well *below* the spectral content of the noise and close to the frequency of the signal. This outcome is consistent with the hypothesis that the signal was being masked by aural distortion products, an interpretation that will be discussed in detail after other data are presented.

When the frequency of the signal was 450 Hz [Fig. 1, panel (b)], the data were best fit by a cosine function having a frequency of 469 Hz (solid line). That frequency accounted for 87% of the variance in the data. Although the spectral content of the noise was the same as when the frequency of the signal was 525 Hz [panel (a)], the frequency that best fits the data is now 71 Hz lower than before. As before, the best fitting frequency is well *below* the spectral content of the noise and close to the frequency of the signal. Overall, based on the frequencies that best fit the data, the patterning of the data in panels (a) and (b) of Fig. 1 is consistent with the hypothesis that the signals were masked by aural distortion products.

Let us now, using the same fitting methods as before, consider how well the data are fit by other frequencies besides the best fitting ones. This will provide an indication of how robust the data are in terms of being characterized by the single cosine functions shown in panels (a) and (b) of Fig. 1. Panel (a) of Fig. 2 shows the percentages of variance accounted for in the data presented in panels (a) and (b) of Fig. 1 as a function of the frequency used to fit the data. Note that the frequencies which account for the largest percentages in the data are clearly different for the two signal frequencies, despite the fact that the external or physical noise was the same in the two conditions. Also note that specification of a *spectral region* rather than a particular frequency appears to be a more accurate description of the quality of fits to the data by single cosine functions. Essentially identical amounts of variance were accounted for by a small band of frequencies surrounding the best fitting one. Still, it is clearly the case that the spectral regions that provide the best fits to the data are those that were expected to provide on-frequency masking due to aural distortion products.

Let us now turn to the data obtained with the lower-frequency band of noise [Fig. 1, panel (c)]. The reader is reminded that the frequency of the signal (525 Hz) in this condition was *above* the spectral content of the noise, which ranged from 0 to 450 Hz.

The behavioral thresholds in panel (c) vary as a function of ITD, indicating that interaurally delaying a noise can differentially affect the detectability of a signal placed spectrally above the highest frequencies in the noise. The maximum release from masking, about 8 dB, occurred when the ITD was 780 μs. The solid line in Fig. 1, panel (c), represents the best fitting cosine function which, in this case, had a frequency of 479 Hz and accounted for 79% of the variance in the data. A frequency of 479 Hz is 29 Hz above the highest components of the noise and 46 Hz below the frequency of the signal.

Panel (b) of Fig. 2 shows the percentages of variance accounted for the data in panel (c) of Fig. 1 as a function of the frequency used to fit the data. Frequencies in the region between about 450 and 510 Hz account for between 72% and 79% of the variance of the data. In this case, the best fitting cosine functions are well *below* the signal frequency and slightly above the highest components of the noise. This outcome, taken in conjunction with the data discussed earlier, suggests that different processes or mechanisms may mediate

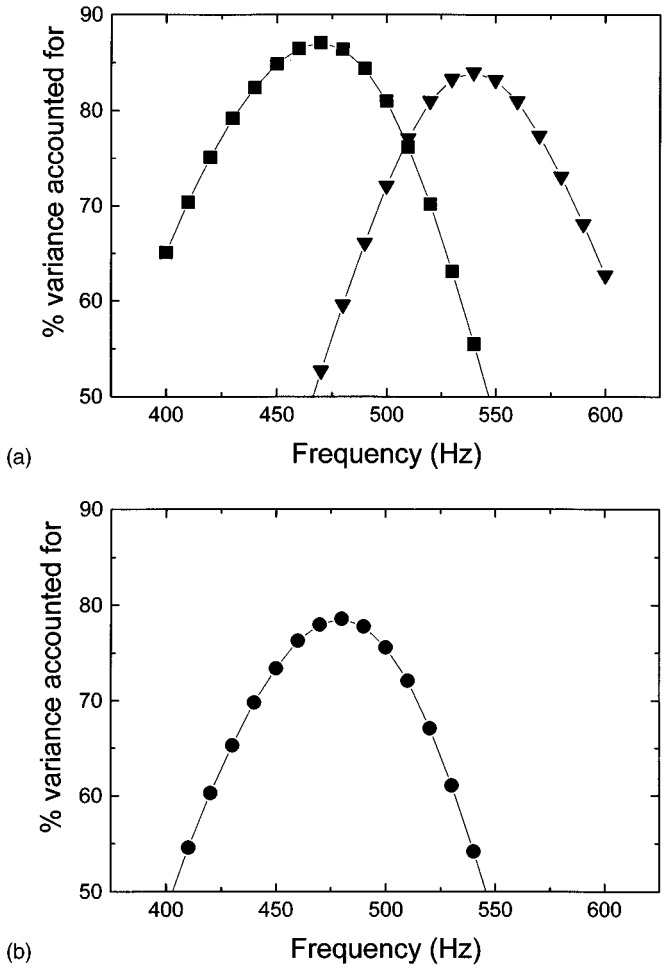


FIG. 2. The percentage of variance in the data of Fig. 1 accounted for by fitted cosines as a function of the frequency of the fitting cosine functions. In panel (a), this value is shown for the thresholds obtained when the spectral content of the noise ranged from 600 to 1100 Hz. The two different curves represent the two different signal frequencies: 450 Hz (closed squares) and 525 Hz (closed triangles). In panel (b), the percentage of variance accounted for is shown for the thresholds obtained when the spectral content of the noise ranged from 0 to 450 Hz and the signal frequency was 525 Hz (closed circles).

binaural detection depending on whether the frequency of the signal is below or above the spectral content of the external noise.

Although the data presented in the three panels of Fig. 1 are well fit by cosine functions, close visual inspection indicates that there are systematic departures from the best fitting cosine functions, especially for the thresholds obtained with the smaller values of ITD. Among the most striking examples are the thresholds in Fig. 1, panel (c), for TDs ranging from 260 to 780 μs. Those data clearly suggest a much more rapid decline in threshold than that described by the best fitting function. We were also concerned that those data appear to have a second maximum at an ITD of about 2210 μs that is located beyond the maximum displayed by the best fitting cosine function of 479 Hz. These aspects of the data are addressed next in Sec. III.

III. DISCUSSION

We begin by discussing the results in terms of Durlach's equalization-cancellation model (Durlach, 1963, 1972). In an NτSo condition such as those investigated here, equalization is accomplished via internal delays that compensate for the ITD in the noise. The stimulus condition is thereby transformed into a new configuration in which the noise is, ideally, interaurally identical (No). After equalization, the signal contains a time-delay equal to the internal delay required to equalize the noise. In this manner, the NτSo condition becomes, after equalization, effectively equivalent to No Sτ, where τ is the time delay of the signal corresponding to the internal delay required to match the external delay of the noise.

Now, in our experiment, if the listeners were able to match (internally) the ITD of the noise, then one would expect that NτSo thresholds would vary in the same manner as would thresholds obtained in an No Sτ condition with the same signal frequency. In order to see if this were true, we retested the same listeners in the three main conditions of our original experiment utilizing an No Sτ stimulus configuration. For each signal frequency, we used several ITDs that were equivalent to phase shifts of up to 180 deg. In addition, for reasons discussed later, we also obtained data using an NπSo stimulus condition.

The new data, obtained in the No Sτ condition, are plotted in the three panels of Fig. 3 as open symbols. The data represented by closed symbols are replotted from Fig. 1.

When the signal was spectrally below the noise [panels (a) and (b) of Fig. 3], thresholds obtained with ITDs of the signal greatly overlap thresholds obtained with ITDs of the noise, provided that the ITD was less than 500 μs or so. For larger values of ITD, thresholds obtained with ITDs of the signal are consistently *lower* than thresholds obtained with ITDs of the noise. Panel (c) of Fig. 3 contains similar trends, save for the fact that the thresholds obtained with ITDs of the signal and thresholds obtained with ITDs of the noise now overlap for ITDs of up to about 800 μs.

We believe that the overlapping of the thresholds in the NτSo and No Sτ conditions is evidence that the external delays were equalized or matched by appropriate internal delays. This statement holds for the noise components originating externally and for noise components generated internally by auditory nonlinearities. In contrast, the nonoverlapping of NτSo and No Sτ thresholds for larger delays (delays between 500 and 1000 μs or so) is taken as evidence that the external delays of the noise *were not matched* internally. This suggests that there is a *limitation* on the size of the ITD that can be matched internally. If that were true, then how does one explain the pattern of thresholds obtained with ITDs too large to be matched internally? Our hypothesis is that, for ITDs too large to be matched internally, detection is based on the interaural correlation coefficient. The interaural correlation coefficient is the value of the interaural correlation function at lag zero. Therefore, decisions based on it do not require equalization accomplished via the delay line. Our hypothesis differs from that made by Durlach (1972) and others, who assume that detection depends on the ''best available internal delay'' independent of whether that delay completely equalizes the external delay.

This line of argument leads us to look for two distinct regions in the data. In the region of ''small'' ITDs (those that

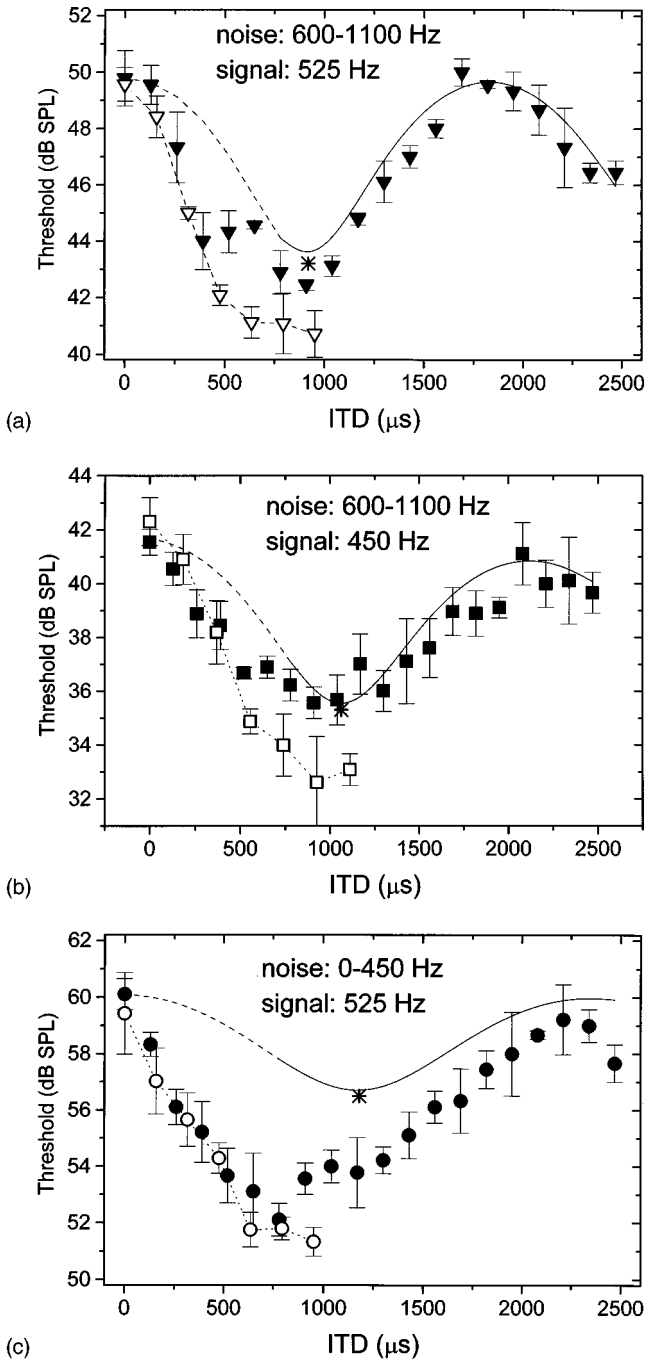


FIG. 3. Open symbols in panels (a)–(c) indicate NoSτ thresholds as a function of the ITD of the signal. Closed symbols in panels (a)–(c) indicate thresholds obtained in the corresponding spectral conditions using the NτSo configuration (these data are replotted from Fig. 1). Different panels correspond to different spectral conditions: the data in panels (a) and (b) were obtained with the spectral content of the noise ranging from 600 to 1100 Hz (and with signal frequencies of 525 and 450 Hz, respectively) and the data in panel (c) were obtained with the spectral content of the noise ranging from 0 to 450 Hz (and a signal frequency of 525 Hz). Solid lines indicate predictions based on the interaural correlation of an interaurally delayed noise stimulus for ITDs greater than 700 μs. The dashed lines represent extensions of the predictions to include ITDs less than 700 μs. NπSo thresholds are indicated by asterisks.

can be matched internally), the pattern of thresholds would reflect the frequency of the signal. In the region of ''large'' ITDs (those that *cannot* be matched internally), the pattern of thresholds would reflect the periodicity in the autocorrelation of the noise that is responsible for the masking.

We now evaluate whether NτSo thresholds obtained with ITDs presumably too large to be matched internally can be explained using the interaural correlation coefficient. We will begin with the data obtained when the frequency of the signal was *below* the spectral content of the noise. We assume that masking in this case was due to ''on-frequency'' aural distortion products and that the patterning of the data will reflect this. The methods used to evaluate this assumption were a bit complicated. We inspected the autocorrelation function of a 100-Hz-wide band of noise centered on the frequency the signal. The bandwidth of 100 Hz was chosen to approximate the bandwidth of noise at the output of a critical band at these low signal frequencies. The spectrum of the bands of noise contained a tilt (a spectral slope of 0.11 dB/Hz) favoring the higher frequencies. That value of spectral slope was determined by taking the 8-dB difference in detection threshold (obtained with an ITD of 0 μs) for the two signal frequencies placed spectrally below the noise and dividing that difference in threshold by the 75-Hz difference in the frequency of the signals. We made the reasonable assumption that the 8-dB difference in diotic (NoSo) thresholds for the two signal frequencies occurred because the noise power responsible for the masking differed by 8 dB across the two signal frequencies. This type of spectrum, ''tilted'' toward the higher frequencies, is consistent with the literature concerning aural distortion products (e.g., Goldstein, 1967; van der Heijden and Kohlrausch, 1995). Because of the spectral tilt, the autocorrelation function of the 100-Hz-wide band of noise, although centered on the frequency of the signal, is quasi-periodic with a frequency slightly *higher* than the frequency of the signal.

We evaluated the autocorrelation functions of the two noises for ITDs greater than 700 μs, a value chosen by considering all the data. The process involved recasting NτSo conditions into NρSo conditions, with ρ being the interaural correlation of noise stemming from aural distortion products. In order to do so, we had to obtain additional data using an NπSo stimulus configuration. These NπSo ($\rho=-1$) thresholds, in combination with the NoSo ($\rho=+1$) thresholds obtained in the main experiment, served as endpoints in the function relating NρSo detection thresholds to ρ. In order to obtain predictions of thresholds for intermediate values of ρ, we utilized the interpolation method recently described by van der Heijden and Trahiotis (1996).

The solid lines in panels (a) and (b) of Fig. 3 represent predictions obtained for ITDs $\geqslant$700 μs using the interaural correlation *coefficient* of the tilted-spectrum bands of noise representing aural distortion products as discussed earlier. The reader is reminded that these predictions were made assuming that external ITDs were *not matched* internally. The correspondence between the predictions and the data appears to be quite good in both cases. When the signal frequency was 525 Hz [panel (a)], the amount of variance in the data accounted for by the predictions was 87%. When the signal frequency was 450 Hz [panel (b)], the amount of variance in the data accounted for by the predictions was 75%.

A corollary of our theoretical position is that extending

the predictions to ITDs *less* than 700 μs (the ITDs we have argued *are* internally matched) would reduce the amount of variance accounted for in the data. Indeed, as shown by the dashed line extensions to our predictions in Fig. 3, the predictions for the smaller ITDs would be quite poor. When we included those values of ITD in the analysis, the amount of variance accounted for decreased and was 67% and 65%, for the 525-Hz and the 450-Hz signals, respectively.

Overall, our analyses strongly suggest that aural distortion products are responsible for the masking that occurred when the frequency of the signal was *below* the spectral content of the noise. In our judgment, the predictions based on aural distortion products are remarkably good, especially when it is considered that the predictions were made by taking into account: only two behavioral thresholds and logical arguments concerning the properties of the aural distortion products assumed to be responsible for the masking.

As a check on the sensitivity of our analysis, we also investigated the autocorrelation function assuming a critical bandwidth of 200 Hz rather that 100 Hz. The amount of variance accounted for when the frequency of the signal was 525 Hz dropped to 72%. The amount of variance accounted for when the frequency of the signal was 450 Hz dropped to 61%. We take this as evidence that the first analysis, using the 100-Hz-wide critical bandwidth, was not a fortuitous success and that our data and methods have sufficient precision.

We now consider predictions for the case when the 525-Hz signal was spectrally *above* the components of the physical noise which ranged from 0 to 450 Hz. We investigated the interaural crosscorrelation of a 50-Hz-wide, rectangular, band of noise centered at 425 Hz. This band of noise was chosen because we assumed that masking was produced by frequencies at the upper edge of the noise and that such a band of noise would not be removed by an internal filter centered on 525 Hz, the signal frequency.

Predictions obtained using the general method described above are shown as the solid line in panel (c) of Fig. 3. Note that, in this case, the predicted thresholds are a poor fit to the data and are consistently *higher* than the obtained thresholds. In fact, using the mean as a predictor would provide a better fit to the data.

It is not the case that our assumption concerning which *frequencies* produced masking is responsible for the poor predictions obtained in this condition. On the contrary, the data obtained with ITDs greater than 700 μs are extremely well fit by single cosine functions having a frequency *within the upper region* of the band of noise. In fact, cosine functions having a frequency in the range of 420–450 Hz, the uppermost region of the noise, each accounted for more than 90% of the variance in the behavioral data. At the same time, a cosine function of 525 Hz, the frequency of the signal, accounted for less than 60% of the variance in the data collected with ITDs greater than 700 μs. Therefore, the inability to predict thresholds when the frequency of the signal was *above* the spectral content of the noise is *not* due to the periodicities in the data being somehow incompatible with the spectral region assumed to be responsible for the masking.

Our interpretation of the data is predicated on cancellation of noise for the smaller delays that can be internally matched and on the use of the correlation coefficient (i.e., the correlation function at lag zero) for larger delays that cannot be internally matched. Of course, it is possible that binaural detection for ''in-between'' values of external delay may depend upon a combination of both modes of processing and/or a differential weighting of internal delays depending on their usefulness. In addition, our interpretations do not explicitly consider the potential use of ''slipped cycles'' for detection when external delays are large enough so that complete cancellation is impossible. As discussed by Rabiner *et al.* (1966), especially on pages 69 and 70, explanations of binaural detection data depending upon the use of slipped-cycle information involve a myriad of complex issues and data. At this time, we cannot provide an analysis of our data that incorporates, let alone resolves, the complex issues discussed by Rabiner *et al.* However, we believe that future data obtained utilizing very large delays (representing multiples of the periods of the signals) in our paradigm with masking produced by aural distortion products may permit the use of information in slipped cycles to be evaluated. Still, it appears to us that our analyses are an appropriate beginning toward the understanding of binaural detection data obtained under conditions where aural distortion products could be responsible for the masking.

We now present arguments that indicate that there is some other, unknown, factor that plays a role in binaural detection when the signal frequency is above the masking noise. Recall that, in order to make predictions for several values of interaural correlation via our interpolation method, we had to measure thresholds using an NπSo configuration. This was done for all three of our main experimental conditions. Those thresholds, which are represented by asterisks in panels (a)–(c) of Fig. 3, were *very close* to their NτSo counterparts when the frequency of the signals was *lower* than the spectral content of the noise [Fig. 3, panels (a) and (b)].

Note, however, that the NπSo threshold measured with the frequency of the signal *above* the spectral content of the masker [Fig. 3, panel (c)], is 3 dB *higher* than its NτSo counterpart. As a consequence, when the frequency of the signal was above the spectral content of the noise, the MLD between NπSo and NoSo conditions is slightly less than 4 dB. In contrast, the MLD between the equivalent NτSo conditions (i.e., an ITD of about 1000 μs) and the NoSo conditions is about 7 dB. This is a very important difference. *A priori*, one would expect that NπSo *thresholds*, for which the interaural correlation of the noise is -1, would be equal to, or lower than, thresholds obtained with any other values of interaural correlation of the masker. This means that the variations in the thresholds obtained when the frequency of the signal was *above* the spectrum of the noise cannot be explained solely in terms of the interaural correlation of the masker. This interpretation comes from relations among the data and does not entail assumptions concerning which spectral components contributed to the masking.

At this time, we have no explanation for the masking that occurred when the signal was spectrally above the noise and when the ITD of the noise was presumably too large to be equalized. In future investigations, we plan to determine

whether, and to what degree, binaural interference effects and binaural masking effects may combine or interact to determine detectability for signals placed spectrally above maskers and/or interferers. Binaural interference refers to degradations in the ability to detect or discriminate interaural differences in cases where masking effects can be ruled out. As discussed by Bernstein and Trahiotis (1995), binaural interference effects are often asymmetric in that low-frequency interferers affect high-frequency targets more than high-frequency interferers affect low-frequency targets. At this time, it appears not too far-fetched to speculate that noises containing large ITDs cannot be equalized and, therefore, cannot be canceled. Consequently, they may remain to affect detectability via binaural interference.

In summary, we have presented data suggesting that aural distortion products that are evoked by a band of noise have the ability to produce binaural masking of tonal signals having a frequency *below* the spectral content of the noise. The detectability of such signals is well accounted for by considering the expected interaural correlation of aural distortion products evoked by interaurally delayed noise. We conclude that aural distortion products can produce binaural as well monaural masking effects (e.g., van der Heijden and Kohlrausch, 1995). On the other hand, data obtained with tonal signals having a frequency *above* the spectral content of the noise are not satisfactorily accounted for by considering the interaural correlation of the noise. It appears that another factor, perhaps binaural interference, limits binaural detectability for signals placed spectrally above masking noise.

ACKNOWLEDGMENTS

The authors thank Dr. Les Bernstein for his careful reading of several drafts of this manuscript. His many suggestions are appreciated. We also thank Dr. H. Steven Colburn, an identified reviewer, who correctly, yet supportively, stated that our original discussion section needed severe reorganization. Supported by research Grant No. NIH DC-00234 from the National Institute on Deafness and Other Communication Disorders, National Institute of Health.

Bernstein, L. R., and Trahiotis, C. (**1995**). ''Binaural spectral interference in detection and discrimination paradigms,'' in *Advances in Hearing Research: Proceedings of the 10th International Symposium on Hearing*, edited by G. A. Manley, G. M. Klump, C. Koppl, H. Fastl, and H. Oekinghaus (World Scientific, Singapore).

Bilger, R. C., and Hirsch, I. J. (**1956**). ''Masking of tones by bands of noise,'' J. Acoust. Soc. Am. **28**, 623–630.

Durlach, N. I. (**1963**). ''Equalization and cancellation theory of binaural masking-level differences,'' J. Acoust. Soc. Am. **35**, 1206–1218.

Durlach, N. I. (**1972**). ''Binaural signal detection: Equalization and cancellation theory,'' in *Foundations of Modern Auditory Theory, Volume II*, edited by J. V. Tobias (Academic, New York), pp. 365–466.

Goldstein, J. L. (**1967**). ''Aural combination tones,'' J. Acoust. Soc. Am. **41**, 676–689.

Hall, J. W., Tyler, R. S., and Fernandez, M. A. (**1983**). ''Monaural and binaural auditory frequency selectivity resolution measured using band-limited noise and notched-noise masking,'' J. Acoust. Soc. Am. **73**, 894–898.

Langford, T. L., and Jeffress, L. A. (**1964**). ''Effect of noise crosscorrelation on binaural signal detection,'' J. Acoust. Soc. Am. **36**, 1455–1458.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

McFadden, D., Russell, W. E., and Pulliam, K. A. (**1972**). ''Monaural and binaural masking patterns for a low-frequency tone,'' J. Acoust. Soc. Am. **51**, 534–543.

Rabiner, L. R., Laurence, C. L., and Durlach, N. I. (**1966**). ''Further results on binaural unmasking and the EC model,'' J. Acoust. Soc. Am. **40**, 62–70.

van der Heijden, M., and Kohlrausch, A. (**1995**). ''The role of distortion products in masking by single bands of noise,'' J. Acoust. Soc. Am. **98**, 3125–3134.

van der Heijden, M., and Trahiotis, C. (**1996**). ''A new way to account for binaural detection as a function of interaural noise correlation,'' J. Acoust. Soc. Am. (accepted for publication).

Cross-spectral and temporal factors in the precedence effect: Discrimination suppression of the lag sound in free-field[a)]

Xuefeng Yang and D. Wesley Grantham
Department of Hearing and Speech Sciences, Vanderbilt Bill Wilkerson Center for Otolaryngology and Communication Sciences, Vanderbilt University Medical Center, Nashville, Tennessee 37232

(Received 14 November 1996; revised 26 February 1997; accepted 5 June 1997)

In an anechoic chamber, subjects were required to discriminate a 20° azimuthal change in a lag sound's position in the presence of a lead sound coming from a different direction. Delay between lead and lag sounds was adaptively varied in several conditions to track discrimination suppression thresholds. In experiment 1, lead and lag stimuli were 5-ms, 1-octave, A-weighted noise bursts (65 dB), with lead and lag parametrically set to center frequencies of 0.5, 2.0, or 3.0 kHz. Discrimination suppression thresholds were higher when lead and lag center frequencies coincided (mean: 11.3 ms) than when they did not coincide (mean: 2.9 ms). These results support the ''spectral overlap hypothesis'' of Blauert and Divenyi [Acustica **66**, 267–274 (1988)], but not the ''localization strength hypothesis'' later proposed by Divenyi [J. Acoust. Soc. Am. **91**, 1078–1084 (1992)]. Spectral overlap and localization strength appear to be two relatively independent factors governing discrimination suppression. It is proposed here that localization strength is weighted more when stimuli are presented via headphones and the only cue to lateral position is the interaural temporal difference, while spectral overlap is weighted more for free-field presented stimuli. In experiment 2, lead and lag stimuli were 8-ms, 1.5-kHz A-weighted tone bursts (65 dB), with lead and lag rise times parametrically set to 0, 2, or 4 ms. In this case the amount of discrimination suppression increased as lead rise time became more abrupt or as lag rise time became more gradual. These results support the localization strength hypothesis: The greater the localization strength of the lead stimulus (independently assessed by measuring its minimum audible angle in isolation), the greater suppression it exerted on discriminability of the lag sound's position. It appears that for stimuli presented in the free-field, spectral overlap is the primary factor affecting discrimination suppression, but when overlap is held constant, abruptness of stimulus onsets governs the amount of suppression. [S0001-4966(97)05510-0]

PACS numbers: 43.66.Pn, 43.66.Qp [RHD]

INTRODUCTION

A. The precedence effect

In a normal room a person is typically exposed not only to the direct sound emanating from a source (e.g., a talker), but also to the reflections of that sound source coming from the walls, ceiling, and floor. Such reflections are not generally noticed by a listener, nor do they substantially disrupt a person's ability to localize the primary sound source, despite generating complex waveforms at a person's two ears that, when integrated over time, do not preserve the normal interaural and spectral cues employed to localize sounds (Wallach *et al.*, 1949; Blauert, 1983). The auditory system is somehow able to give precedence to the first-arriving directional information (i.e., from the direct sound) and to discount or suppress later-arriving information (from the reflections). This phenomenon is called the ''precedence effect'' (Wallach *et al.*, 1949; Zurek, 1987).

The precedence effect is typically investigated in the laboratory with a simplified arrangement consisting of a lead sound coming from one direction and a single simulated reflection (the lag sound) coming after a certain delay from a different direction. Within this framework, three aspects of the precedence effect have been investigated as a function of the delay between lead and lag sounds (i.e., the *echo delay*):

(1) echo suppression, or fusion: are the lead and lag sounds perceived as a single composite event, or as two separate events? Experiments in this category have been reported by Blauert (1983; pp. 222–235), Clifton and Freyman (1989), Freyman *et al.* (1991), and Yang and Grantham (1997).
(2) localization dominance: If a single, fused event is perceived and the subject is asked to localize it, to what extent does s/he point correctly to the direction of the lead sound? The ''adjustment'' experiments of Wallach *et al.* (1949), Yost and Soderquist (1984), and Shinn-Cunningham *et al.* (1993, 1995) fall into this category, as well as the pointing experiment reported by Hafter *et al.* (1992).
(3) lag discrimination: To what extent can small directional changes in the lag sound be discriminated? Many experiments have been conducted in this category, both for precedence-simulation under headphones (Zurek, 1980; Saberi and Perrott, 1990) and in the free-field (Perrott

[a)]This paper represents a portion of the dissertation completed by the first author in partial fulfillment of the requirements for the doctoral degree in the Division of Hearing and Speech Sciences at Vanderbilt University, 1995. Some of the results were reported at the 129th meeting of the Acoustical Society of America, May 1995, Washington, DC [J. Acoust. Soc. Am. **97**, 3280 (A) (1995)].

et al., 1989; Freyman *et al.*, 1991; Hafter *et al.*, 1992; Litovsky and Macmillan, 1994; Grantham, 1996).[1]

The present investigation is concerned with the third of these aspects of precedence—lag discrimination. Since the discrimination of the lag sound's position is typically worse in the presence of the lead sound than when it is absent, we will refer to the *suppression* of lag discrimination produced by the lead sound, or, more briefly, to *discrimination suppression.*

B. Discrimination suppression of ''nonechoes''

Under normal circumstances, a direct sound and its echo are highly correlated sounds; that is, they are spectrally and temporally similar. Most of the previously cited studies of discrimination suppression have, in fact, employed lag stimuli that could be considered to be ''real echoes'' of the lead sound. However, some recent studies have investigated discrimination suppression in the more general situation involving *any* two sounds presented in close temporal proximity. Divenyi (1992) has referred to this more generalized task as ''suppression of nonechoes.''

Divenyi and Blauert (1987) and Blauert and Divenyi (1988) investigated discrimination suppression under headphones by measuring the interaural time difference (ITD) threshold for a lag sound in the presence of a diotic lead sound. The stimuli were clicks or noise bursts 2–10 ms in duration, and echo delay was either 3 or 6 ms; lead and lag bursts were independently filtered in various ways. These investigators found that the most discrimination suppression (i.e., the greatest elevation of ITD threshold) occurred when there was complete spectral overlap between the lead and lag sounds, weaker suppression occurred when the frequency content of the lag sound was partially or completely above that of the lead sound, and no suppression occurred when the frequency content of the lag was below that of the lead. Blauert and Divenyi proposed that, like other auditory phenomena, discrimination suppression operates within frequency bands. That is, the magnitude of suppression is directly related to the amount of spectral overlap between the lead and lag sounds after processing by the peripheral auditory filters. They also showed that discrimination suppression was not a simple consequence of forward masking: in all cases the lag sound was clearly *detectable*, even when its ITD threshold was too large to be measured.

Figure 1 presents a simplified picture of the predictions of a ''spectral overlap hypothesis,'' where the amount of discrimination suppression is plotted as a function of the lead sound center frequency, with the lag sound center frequency as the parameter. In this idealized schematic, significant suppression occurs only when the lead and lag sounds have identical power spectra. In general, the results of Divenyi and Blauert (1987) and Blauert and Divenyi (1988) conform to this hypothesis.

C. Divenyi's localization strength hypothesis

Divenyi (1992) conducted an experiment very similar to those reported by Blauert and Divenyi (1988), but obtained results that differed in one important respect from the earlier

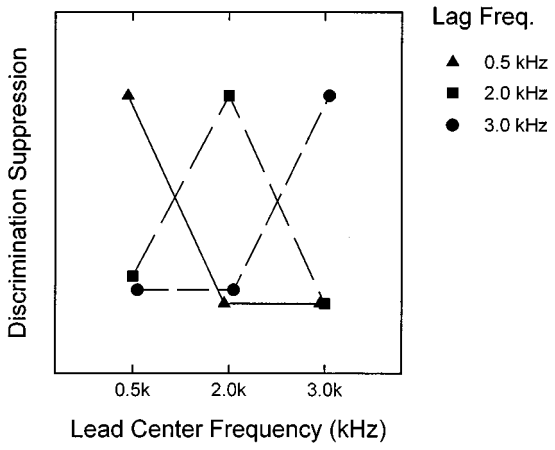


FIG. 1. Predicted amount of discrimination suppression according to ''spectral overlap'' hypothesis: the amount of suppression is determined by the degree of overlap of the magnitude spectra of the lead and lag sounds. In this simplified diagram, maximum suppression occurs only when the center frequencies of the lead and lag stimuli are the same. (Frequency values shown correspond to the center frequencies employed in the present experiment 1.)

findings. Employing a lag stimulus centered at 2.0 kHz, Divenyi found *more* discrimination suppression when the lead stimulus was centered at 0.5 kHz than when it was centered at the same frequency (2.0 kHz) as the lag. The low-frequency stimulus appeared to be dominant in terms of its ability to suppress discrimination of the lag position of the high-frequency stimulus. This result clearly does not fit the explanation for discrimination suppression based on spectral overlap (Blauert and Divenyi, 1988).

Divenyi speculated that the differences between his later (1992) results and the earlier results he had obtained in collaboration with Blauert may have been due to stimulus differences (e.g., narrower bandwidths and a different selection of center frequencies in the later study), though it was not clear *why* such stimulus differences led to the different results.[2] In any case, Divenyi (1992) concluded that, although spectral overlap may have some influence in discrimination suppression, there must be ''some other, more compelling and hitherto unknown factor that impedes lateralization of the [lag stimulus] when the [lead stimulus] frequency is low, but not when it is high'' (p. 1081). He went on to suggest that one such possible factor is the ''localization strength'' of the lead stimulus, where localization strength refers to the ''perceptual salience of spatial information.''

One way to quantify localization strength for a given stimulus is to measure its ITD threshold in isolation: the lower the ITD threshold, the greater the localization strength. Using this measure, Divenyi (1992) showed in a separate experiment that localization strength of his narrowband stimuli varied inversely with center frequency: for stimuli centered at 0.5, 2.0, and 3.0 kHz the greatest localization strength was associated with 0.5 kHz and the least with 3.0 kHz. Thus his finding in the main experiment that discrimination suppression decreased as lead sound center frequency increased, indicates that suppression was more closely asso-

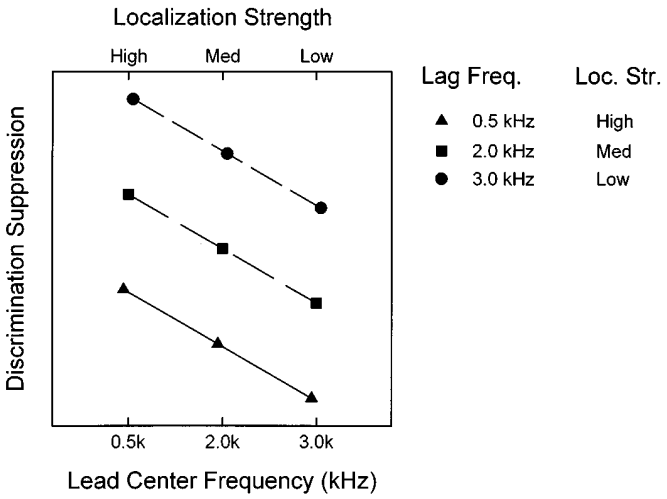


FIG. 2. Predicted amount of suppression according to ''localization strength'' hypothesis: suppression decreases as localization strength of the lead sound decreases and increases as the localization strength of the lag sound decreases. As indicated in the legend and on the upper abscissa, localization strength is assumed to decrease as center frequency increases; this relationship may be established in an independent experiment that measures the minimum audible angle (MAA) of the three stimuli in isolation (see Appendix). (Frequency values shown correspond to the center frequencies employed in the present experiment 1.)

ciated with lead sound localization strength than with spectral overlap between lead and lag sounds.

A more general depiction of the localization strength hypothesis is shown in Fig. 2, where it can be seen that the predicted amount of discrimination suppression is based on the *relative* localization strengths of the lead *and lag* sounds. Since localization strength is a decreasing function of center frequency (based on the ITD threshold measurements made by Divenyi), the abscissa shows localization strength going from high on the left to low on the right (see upper axis). Suppression is expected to increase as lead sound localization strength increases, as pointed out by Divenyi, and as shown by the negative slope of the three curves in the figure. At the same time, suppression is expected to *decrease* as lag sound localization strength increases; that is, as the lag sound becomes more localizable, discriminating changes in its position should become more resistant to suppression by the lead sound. This is shown in the figure by the relative positions of the three curves, with the lowest curve representing the lag sound with the greatest localization strength.

Other reports in the literature have suggested that low-frequency stimuli (i.e., stimuli high in localization strength) can dominate in a precedence effect situation. Scharf (1974) showed that when two tones of different frequencies are presented from two different azimuths in the horizontal plane, they can appear to be coming from the same location. The subjects' directional responses were determined primarily by which of the two tones came on first; but a secondary effect was a consistent bias to respond in the direction of the *lower*-frequency tone. Scharf pointed out that such low-frequency dominance had been reported for lateralization of headphone-presented stimuli as well (e.g., Yost *et al.*, 1971).

Shinn-Cunningham *et al.* (1995) also reported some data that support the localization strength hypothesis. The stimuli employed by these investigators were similar to those employed in the discrimination suppression studies discussed above: a lead noise burst and a lag noise burst were presented over headphones, where the ITD and the frequency content of the lead and lag sounds were independently manipulated. However, the subject's task was not to discriminate the lag's position (as in the earlier studies), but to indicate the apparent lateral position of the composite image inside the head (i.e., this was a ''localization dominance'' experiment). In this case, suppression by the lead sound was quantified by the extent to which the subject's lateral position response corresponded to the ITD of the lead stimulus and was independent of the ITD of the lag stimulus. [In an earlier study Shinn-Cunningham *et al.* (1993) had shown that localization dominance (or ''adjustment tasks'') and lag discrimination tasks can be modeled within a single framework that quantifies the weights assigned to the lead and lag sounds.]

In agreement with the localization strength hypothesis, Shinn-Cunningham *et al.* (1995) found a large amount of suppression in the ''$L-H$'' condition (i.e., when the lead sound was a low-frequency noise and the lag was a high-frequency noise), less suppression in the ''$L-L$'' condition, and the least amount in the ''$H-L$'' condition. Little suppression in the latter two conditions is predicted by the localization strength hypothesis, since the low-frequency lag stimulus (with high localization strength) is expected to *resist* suppression. They also found a large amount of suppression in the ''$H-H$'' condition, which does not necessarily contradict the localization strength hypothesis, but is different from Divenyi's finding of little or no suppression when lead and lag were both high-frequency stimuli.[2] Taken together, these data support the localization strength hypothesis and do *not* support the spectral overlap hypothesis.

D. Rationale for the present experiments

The present experiments were undertaken in an attempt to determine which (if either) of the two hypotheses represented in Figs. 1 and 2 would better predict discrimination suppression in a free-field situation. We employed a variant of the usual method of measuring discrimination suppression. In measuring lag sound discriminability most previous studies have held echo delay constant and varied the angular separation between two potential lag events to measure a minimum audible angle (MAA) (or, under headphones, an ITD threshold) for the lag source (Zurek, 1980; Divenyi and Blauert, 1987; Blauert and Divenyi, 1988; Perrott *et al.*, 1989; Saberi and Perrott, 1990; Divenyi, 1992; Hafter *et al.*, 1992; Litovsky and Macmillan, 1994). Here, however, we held the angular distance between two lag positions constant (at 20°) and varied the *echo delay* to obtain a critical delay at which the angular discrimination could just be made (see Freyman *et al.*, 1991; Clifton *et al.*, 1994; Grantham, 1996). Thus, the dependent measure quantifying lag discriminability is a ''discrimination suppression threshold'' measured in milliseconds of echo delay.

In experiment 1, octave-band random noise signals with center frequencies from 0.5 to 3.0 kHz were employed (similar to those employed by Blauert and Divenyi, 1988). The

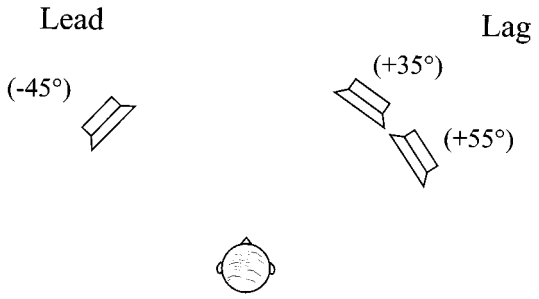


FIG. 3. Loudspeaker arrangement for the measurement of discrimination suppression threshold. On each trial the lead stimulus was presented from the loudspeaker at −45°, and the lag stimulus, presented after a variable delay, was presented from one of the two loudspeakers on the subject's right side. Subject's task was to indicate which of the two loudspeakers on the right emitted the lag sound.

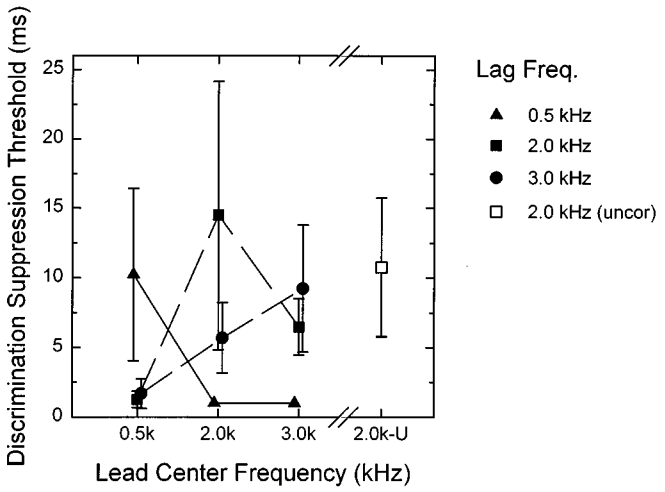


FIG. 4. Mean discrimination suppression thresholds across the five subjects in experiment 1, plotted as a function of lead sound center frequency. Parameter is the lag sound center frequency. The open square plotted to the right (at 2.0 k-U) represents threshold in the case when the lead and lag sounds were both centered at 2.0 kHz, but the waveforms were uncorrelated. Error bars show ±1 standard deviation around the means.

results were consistent with the spectral overlap hypothesis and not the localization strength hypothesis. In experiment 2, frequency of a pure-tone stimulus was held constant (at 1.5 kHz) and localization strength was manipulated by varying the stimulus rise time. In this case, the data were consistent with the localization strength hypothesis.

I. EXPERIMENT 1

A. Method

1. Subjects

Five adults (two males, three females; aged 23 to 38 years) with normal hearing in both ears were recruited to participate in this study. All subjects were paid and none was familiar with the purposes and hypotheses of the study. Subjects were tested individually in sessions that lasted one to two hours. Frequent breaks were given during each session.

The first two–four sessions for each subject were considered practice, and the data from these sessions were not included in the later analyses; subsequent inspection of the data indicated that performance had stabilized by this time.

2. Apparatus and stimuli

The experiment was conducted in a $6\times6\times6$-m anechoic chamber lined on all six surfaces with 0.75-m foam wedges. During a session the subject was seated in a chair at the center of the chamber. The subject's head was not restrained, but s/he was instructed to face directly ahead and not move the head during the run. A semicircular array of loudspeakers (JBL 8110H high-compliance full-range) was positioned in front of the subject at ear level, 1.8 m distant. From this array three loudspeakers were employed, located at azimuths of −45°, +35°, and +55° (see Fig. 3). The room lights were extinguished in order to eliminate possible visual influences on performance.

Stimuli were 5-ms bursts of Gaussian noise, output at a 20-kHz rate through a two-channel 12-bit DAC, and filtered to be 1-octave wide at a center frequency of 0.5, 2.0, or 3.0 kHz (see footnote 3). Bursts were always presented in pairs—one from the lead loudspeaker at −45° and one from one of the two lag loudspeakers at +35° or +55°. Each new burst pair was unique, generated from a Gaussian random number generator at run time. Level of the bursts (calibrated separately for each loudspeaker) was set such that the average A-weighted intensity was 65 dB when played continuously and measured at the position of the subject's head.

In different conditions the lead and lag bursts were each set to one of the three center frequencies, yielding a total of nine conditions. The main set of nine conditions included three in which the lead and lag were centered at the same frequency; in these cases the temporal waveforms were identical for the pair within each trial (correlated waveforms). A tenth condition consisted of a 2.0-kHz lead sound paired with a 2.0-kHz lag sound, where the two waveforms were *un*correlated (i.e., average correlation=0). Discrimination suppression thresholds were measured in each of the ten conditions.

3. Procedure

On each trial, a single lead–lag burst pair was presented with a specified echo delay, and the subject's task was to choose which of the two lag loudspeakers was activated, using a lap-held response box. Immediate feedback was given by LEDs on the response box. Echo delay was varied adaptively to track threshold. Each adaptive block began with an initial echo delay of 60 ms between the lead and lag sounds (a condition subjects found easy); after two correct responses in a row, echo delay decreased (by a factor of 1.4), and after each incorrect response, echo delay increased by this factor. The block terminated after ten reversals of echo delay change, and discrimination suppression threshold (70.7% correct discrimination level) was computed as the mean of the last eight of these reversals (Levitt, 1971). If ten reversals had not occurred within 50 trials, the block terminated and the run was discarded (this rarely happened).

At least four thresholds (i.e., four runs) were obtained from each subject. If the standard deviation for a run (calculated across reversal points) was greater than one-third of the

threshold magnitude, that run was discarded. If the standard deviation calculated *across the four thresholds* was greater than one third of the average of the four thresholds, two more thresholds were obtained for that condition, and all six were included in the final analysis. Additional threshold runs were obtained in about 50% of the cases.

B. Results and discussion

The mean discrimination suppression thresholds averaged across the five subjects are presented in Fig. 4. Mean thresholds for the ten conditions ranged from 1 to 14 ms (see footnote 4). A two-way analysis of variance was performed in which the two within-subjects factors were *lead sound center frequency* (three levels: 0.5, 2.0, and 3.0 kHz) and *lag sound center frequency* (three levels: 0.5, 2.0, and 3.0 kHz). The interaction between lead sound center frequency and lag sound center frequency was significant [$F(4,16)=10.77$, $p<0.001$].

Follow up analyses revealed that discrimination suppression thresholds were significantly higher in the conditions where the lead sound and lag sound had the same center frequencies than when their center frequencies differed. For example, when the lag sound was centered at 0.5 kHz, the discrimination suppression threshold was significantly higher for a lead sound centered at 0.5 kHz than for a lead sound centered at 2.0 kHz [$F(1,4)=11.18$, $p<0.05$] or 3.0 kHz [$F(1,4)=11.18$, $p<0.05$]. When the lag sound had a center frequency of 2.0 kHz, the discrimination threshold was significantly higher for a lead sound centered at 2.0 or 3.0 kHz than for a lead sound centered at 0.5 kHz [$F(1,4)=10.57$, $p<0.05$; or $F(1,4)=68.33$, $P<0.01$]. When the lag sound was centered at 3.0 kHz, the discrimination suppression threshold for lead sound centered at 3.0 kHz was significantly higher than for lead sound centered at 0.5 kHz [$F(1,4)=17.83$, $p<0.025$] or at 2.0 kHz [$F(1,4)=12.75$, $p<0.025$]. Furthermore, linear contrasts indicated that the thresholds for the three matching (same-spectrum) conditions (mean threshold: 11.3 ms) were significantly different from those of the six nonmatching conditions (mean threshold: 2.9 ms) [$F(1,16)=68.1$, $p<0.001$].

The results are in general agreement with those of Blauert and Divenyi (1988) and support the conclusions from these authors: discrimination suppression operates within frequency channels, and the extent of discrimination suppression is directly related to the amount of spectral overlap between lead and lag sounds after peripheral filtering by the auditory system. Thus, the data support the spectral overlap hypothesis (Fig. 1).

C. Localization strength versus spectral overlap

Localization strength of the three stimuli of experiment 1 was measured for the five subjects at an azimuth of $+45°$ (the angular position corresponding to the average lag source position in the discrimination task). The details of this measurement are described in the Appendix. Briefly, localization strength was determined to be greatest for the 0.5-kHz noise and least for the 3.0-kHz noise (Fig. A1), in agreement with the headphone results of Divenyi (1992). Comparing the discrimination thresholds (Fig. 4) to the predicted results of the localization strength hypothesis (Fig. 2), we can see that relative localization strength had no apparent effect on discrimination performance. Evidently, under the conditions employed in the present experiment, spectral overlap is the dominant factor for discrimination suppression, and localization strength does not play a significant role.

Given the present results, which support the spectral overlap hypothesis, and Divenyi's (1992) results, which support the localization strength hypothesis, it is of interest to examine the differences between the respective experiments to determine under what circumstances each factor affects discrimination suppression. The most obvious difference between the two experiments was the manner of stimulus presentation: Divenyi presented signals over headphones, while we presented signals in free-field. Possibly localization strength is the primary factor determining discrimination suppression when ITD is the sole cue for location (Divenyi, 1992; Shinn-Cunningham *et al.*, 1995). It is well known that low-frequency stimuli can dominate (interfere with) high-frequency ITD detection when the two stimuli are presented *simultaneously* (e.g., Bernstein and Trahiotis, 1995), and perhaps the dominance of low-frequency lead sounds in (headphone) discrimination suppression experiments is related to this binaural interference effect.

On the other hand, it is possible that this low-frequency dominance is less important in free-field situations where interaural level differences (ILDs) and spectral cues may also contribute location information. Our finding that there was virtually no suppression of discrimination of 2.0- or 3.0-kHz lag stimuli in the presence of the 0.5-kHz lead sound indicates that location (or some other) information was available to discriminate changes in the positions of these high-frequency stimuli despite the probable loss of ITD sensitivity. According to this scenario, low-frequency lead stimuli may cause loss of sensitivity to ITD cues, but not to ILD and spectral information. In a free-field situation the dominant factor affecting discrimination suppression appears to be the spectral overlap between lead and lag sounds.

If spectral overlap is the primary factor in free-field situations while localization strength is the primary factor for headphone-presented stimuli, we are left with the puzzle of why Blauert and Divenyi (1988), in headphone experiments, obtained data apparently supporting the spectral overlap hypothesis. Despite this apparent support of the spectral overlap hypothesis, however, a careful inspection of Blauert and Divenyi's data indicates that localization strength may have indeed influenced their subjects' performance. For example, they reported partial suppression in the cases when the lead was low frequency (0.5 kHz) and the lag was high frequency (1.5 or 4.5 kHz), while in comparable conditions in the present experiment we found no suppression. Similarly, they found ''complete suppression'' when the lead was 1.5 kHz and the lag was 4.5 kHz (we found only a modest amount of discrimination suppression for a much closer frequency spacing: lead of 2.0 kHz and a lag of 3.0 kHz). In short, we believe that the amount of ''upward spread of suppression'' found by Blauert and Divenyi was greater than can be accounted for by simply considering the asymmetry of periph-

eral filters, and that it probably reflects the remote spectral effects of localization strength as reported by Divenyi (1992) and Shinn-Cunningham *et al.* (1995).

D. Other possible factors governing the relative influence of spectral overlap and localization strength in precedence-effect experiments

Besides the obvious difference of manner of stimulus presentation between Divenyi's and the present experiments, other stimulus differences may have contributed to the different results. For example, while Divenyi's lead sound was diotic (midline), our lead sound was presented from $-45°$. Possibly the off-midline position of our lead sound resulted in an attenuation of the potential effects of localization strength in our experiment (i.e., MAAs are generally smallest for midline stimuli, implying that they should have the greatest localization strength). Also, Divenyi's stimuli were deterministic FM tones, the durations of which were shorter and the bandwidths of which were narrower than those employed in the present experiment. Such stimulus differences may have interacted with the other stimulus differences mentioned to affect the relative influence of spectral overlap and localization strength in discrimination suppression. Further experiments will be required to map out the exact conditions under which these two factors operate.

Another factor that may affect the relative contributions of spectral overlap and localization strength to discrimination suppression is the type of task employed (see Introduction). While results from different *lag discrimination* experiments have supported both the spectral overlap (Blauert and Divenyi, 1988; present experiment 1) and localization strength hypotheses (Divenyi, 1992; present experiment 2), relevant *localization dominance* experiments to our knowledge have supported only the latter hypothesis (Scharf, 1974; Shinn-Cunningham *et al.*, 1995). Whether localization strength is necessarily always dominant when the subject's task is to localize the composite image will require further investigation.

E. The effect of temporal coherence of lead and lag stimuli

To test whether the waveform correlation between lead and lag sounds (where both were centered at 2.0 kHz) affected the magnitude of discrimination suppression, thresholds were compared for the correlated 2.0-kHz lead/2.0-kHz lag condition (the filled square plotted at 2.0 kHz in Fig. 4) and the uncorrelated 2.0-kHz lead/2.0-kHz lag condition (the open square). Although the mean threshold for the correlated condition (14.5 ms) was somewhat higher than that for the uncorrelated stimuli (10.8 ms), this direction of results occurred for only three of the five subjects, and a t-test revealed no significant difference [$t=1.21$, $p>0.291$]. Thus, the degree of suppression in the case of a 2.0-kHz lead and 2.0-kHz lag signal did not depend on whether the two sounds were correlated or uncorrelated. This finding is in agreement with results from Zurek (1980) and Blauert and Divenyi (1988), who came to the same conclusion with other stimuli.

F. Summary

The results of this experiment suggest that in free-field listening situations the discrimination suppression mechanism(s) operate within frequency bands; i.e., the primary factor affecting the suppression of lag directional information is the degree of overlap between the magnitude spectra of lead and lag stimuli. The correspondence between *the phase* spectra of the lead and lag sounds (i.e., temporal similarity) is apparently of little or no importance for discrimination suppression.

Might localization strength have an effect on free-field discrimination suppression if the dominant factor of spectral content is held constant? To explore this possibility, the next experiment varied localization strength by manipulating a temporal feature of the stimuli (the rise time), while holding center frequency constant.

II. EXPERIMENT 2

In experiment 2 both lead and lag sounds were pure tones with the same frequency (thus, there was spectral overlap between lead and lag sounds in all conditions). In this case, the rise times for lead and lag sounds were independently manipulated. The primary interest was whether the discrimination suppression threshold would be associated with localization strength, as predicted by Divenyi's localization strength hypothesis, or with temporal similarity between the lead and lag sounds (i.e., the correspondence of their respective rise times).

A. Method

1. Subjects and apparatus

The five subjects from experiment 1 participated in experiment 2. Testing was conducted in the same anechoic chamber as previously described.

2. Stimuli

A 1.5-kHz pure tone was used as the stimulus in this experiment (this frequency was chosen because pilot data indicated that it showed the most systematic variation of suppression as a function of rise time). The duration of the pure tone was held constant at 8 ms (this longer duration was selected in order to enable a sufficient range in which to manipulate the rise time). The decay time of the pure tone was held constant at 2 ms, while the rise time was set to one of three values (0, 2, or 4 ms). These rise times were selected based on our pilot data, which indicated that this range of variation was associated with significantly different MAA thresholds (see the Appendix). There were nine conditions, consisting of the factorial combination of the three lead sounds (i.e., having one of three rise times) paired with the three lag sounds. Discrimination suppression thresholds were measured in each condition.

3. Procedure

All instructions and procedural details were the same as those employed in experiment 1: discrimination suppression thresholds were computed in each of the nine conditions as the average of 4–6 adaptive threshold runs, in which thresh-

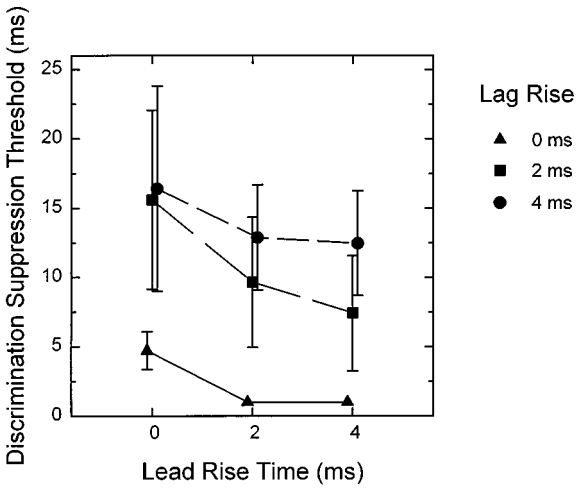


FIG. 5. Mean discrimination suppression thresholds across the five subjects in experiment 2, plotted as a function of lead sound rise time. Parameter is the lag sound rise time. Error bars show ± 1 standard deviation about the means.

old was determined from the last eight of ten reversals in echo delay change. The same rules for inclusion were employed as described for experiment 1.

B. Results and discussion

Figure 5 presents the average discrimination suppression thresholds across the five subjects for all nine conditions. Mean thresholds for the nine conditions varied from 1 to 16 ms. In general, greater suppression was associated with more abrupt *lead* sound rise times and more gradual *lag* sound rise times. A two-way analysis of variance was performed with lead sound rise time (three levels: 0, 2, and 4 ms) and lag sound rise time (three levels: 0, 2, and 4 ms) as within-subjects factors. The results indicate that the interaction between lead sound rise time and lag sound rise time was significant $[F(4,16)=14.11,\ p<0.001]$. The main effects of lead sound rise time $[F(2,8)=27.12,\ p<0.001]$ and lag sound rise time $[F(2,8)=19.00,\ p<0.001]$ were also significant.

Eighteen follow-up tests were performed, consisting of all pairs of conditions in which either the lead or lag sound rise time was held constant (e.g., 0 vs 2-ms lead sound rise time yielded three different tests: one each for lag rise time $=0$, 2, or 4 ms). Of these 18 pairwise tests, 13 showed significant differences. Three of the five nonsignificant tests were obtained in conditions in which the lag rise time was 4 ms (circles in Fig. 5). For such a slow lag rise time, suppression may have reached a ceiling level, such that no manipulation of lead rise time could produce any further suppression.

Localization strength of the three stimuli employed in this experiment (quantified as the MAA) was independently measured at an azimuth of $+45°$ (see the Appendix). As expected based on previous data (Rakerd and Hartmann, 1986), localization strength increased as rise time decreased. Based on this finding, the results in Fig. 5 are in qualitative agreement with Divenyi's localization strength hypothesis: The more abrupt the rise time of the lead stimulus (i.e., the greater its localization strength), the more it suppressed localization information in the lag stimulus. Conversely, the more abrupt the rise time of the lag stimulus (i.e., the greater its localization strength), the more resistant to discrimination suppression it was. The results in Fig. 5 show the same general pattern as the predictions in Fig. 2 (where rise time replaces center frequency as the variable associated with localization strength).

The results from this experiment thus support Divenyi's hypothesis. However, one must bear in mind that the pattern shown in Fig. 5, though demonstrating a relationship between relative localization strength and discrimination suppression, does not demonstrate a causative relationship. It may not be relative localization strength *per se* that determines the magnitude of suppression, but rather some *other* aspect of the rise time. This other aspect of rise time may be the ''true'' determinant of suppression.

1. Spectral differences associated with rise-time differences

One reasonable hypothesis is that spectral features associated with the different rise times are the critical determining factors underlying discrimination suppression. This hypothesis is especially worth investigating in view of the results of experiment 1, which demonstrated the importance of spectral overlap for suppression.

To assess the potential importance of the spectra associated with rise times, magnitude spectra were computed for each of the three tone-bursts employed in experiment 2. These are plotted in Fig. 6 (the curves are offset from each other by 20 dB for clarity). From the figure it is evident that there is considerably greater off-frequency energy (>20 dB) in the tone with 0-ms rise time (top plot) than in the other two tones; it is possible this spectral difference contributed to the results shown in Fig. 5 (especially the resistance to suppression of a lag sound with a 0-ms rise time). Note also that the greater suppression obtained with 0-ms rise-time lead stimuli is consistent with Divenyi's localization strength hypothesis as originally formulated in the frequency domain; that is, to the extent that the greater low-frequency energy for this stimulus determines its localization strength, it would be expected to exert relatively more suppression on the lag sound than would the 2- or 4-ms rise-time lead stimuli.

Spectra of the stimuli with 2- and 4-ms rise times are more similar to each other than either is to that of the 0-ms rise-time stimulus. While there are differences in the shape and pattern of the spectral side lobes of these two stimuli, the overall slope of energy attenuation is the same in the two cases. Given that the first side lobes are already down 15–19 dB *re*: the center lobe, we characterize the spectral magnitude differences between the 2- and 4-ms stimuli as subtle, at best. Since there were some significant differences in performance when rise time changed from 2 to 4 ms (especially for a change in *lag* rise time when the lead rise time was 2 or 4 ms—see Fig. 5), we conclude that magnitude spectra, though perhaps a contributing factor, cannot be the only determinant of suppression.

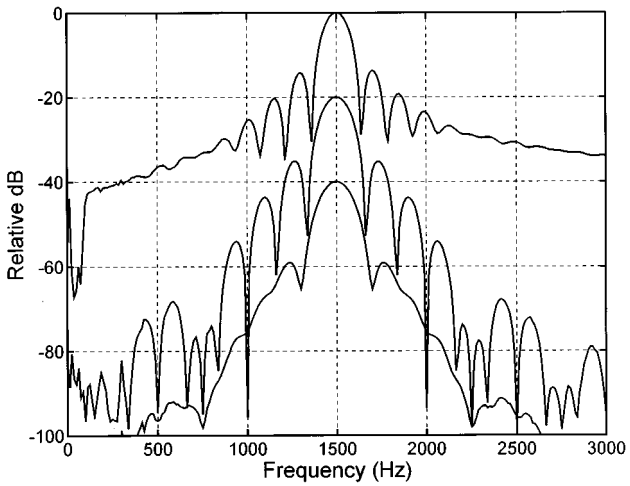


FIG. 6. Magnitude spectra of the three stimuli employed in experiment 2. Spectra were generated by computing an FFT on the entire 8-ms waveform (boxcar window). The three plots are offset from each other by 20 dB for clarity. Upper plot: 0-ms rise time; middle plot: 2-ms rise time; lower plot: 4-ms rise time.

2. Binaural interference

The localization strength hypothesis as well as Divenyi's (1992) and Shinn-Cunningham *et al.*'s (1995) data are reminiscent in some ways of ''binaural interference,'' a phenomenon that has received increased attention in recent years (Woods and Colburn, 1992; Bernstein and Trahiotis, 1995; Heller and Trahiotis, 1995, 1996). Binaural interference may occur when a target is presented *simultaneously* with a second, interfering stimulus at a remote frequency; it is measured as the increase in interaural temporal difference (ITD) threshold of the target in the presence of the interferer, compared to its threshold in isolation. There is considerable binaural interference when the target is a 4.0-kHz band of noise and the interferer is a 500-Hz diotic band of noise, but when the roles of the two stimuli are reversed, there is practically no interference for the ITD threshold of the 500-Hz stimulus (McFadden and Pasanen, 1976; Bernstein and Trahiotis, 1995; Heller and Trahiotis, 1995). As with Divenyi's (1992) data, the amount of ITD threshold elevation (or the amount of lead dominance in Shinn-Cunningham *et al.*'s 1995 experiment) appears to be related to the relative ''localization strengths'' (i.e., the ITD thresholds in isolation) of the target and the interferer. Thus, like Divenyi (1992), Heller and Trahiotis (1995) found that the amount of interference for ITD threshold of a high-frequency stimulus by a low-frequency interferer *increased* as the frequency of the interferer decreased. This result is predicted by the inverse variance model described by Buell and Hafter (1991).

The primary difference in the interference paradigm and that employed by Divenyi (1992) is that the former employed targets and interferers that overlapped temporally, whereas Divenyi's stimuli generally did not overlap in time. Several of the interference experiments have investigated the effect of turning on the interferer prior to the onset of the target (thus producing an onset fringe), though in all cases the interferer remained on throughout the duration of the target. In most cases, the effect of the fringe was to *reduce* the amount of interference (Stellmack, 1994; Bernstein and Trahiotis, 1995; Heller and Trahiotis, 1995). Interestingly, Divenyi found that the condition in which there was some temporal overlap between lead and lag stimuli (for echo delay=2 ms) generally produced less suppression than the nonoverlapping condition (for echo delay=5 ms). It is not clear how the interference results should be compared to those of Divenyi, whose stimuli did not generally overlap in time, but whose durations (4–10 ms) were much shorter than those in the interference experiments (generally greater than 100 ms). If there is a connection between binaural interference and ''suppression of nonechoes,'' understanding the nature of the relationship will require further investigation.

C. Summary

The results from experiment 2 support the conclusion that the onsets (rise times) of the lead and lag sounds play an important role in discrimination suppression when the spectra of the lead and lag sounds are similar. Although the results thus support Divenyi's localization strength hypothesis, they do not allow a determination of whether relative localization strength *per se* is the critical factor underlying discrimination suppression, or whether some other feature(s) associated with rise time are the determinant factors. In any case, temporal similarity between the lead and lag sounds appears not to be an important factor for discrimination suppression, since maximum suppression did not generally occur for cases in which lead and lag signals had identical rise times.

III. SUMMARY AND CONCLUSIONS

Subjects were presented pairs of brief (5–8 ms) stimuli in an anechoic chamber, in which the lead burst came from a loudspeaker at $-45°$ azimuth, and the lag burst came from a loudspeaker at either $+35°$ or $+55°$ azimuth. Echo delay (time between the onset of the lead and lag burst) was varied adaptively to determine the minimum delay at which subjects could just discriminate the direction of the lag burst (70.7% correct). The dependent measure, discrimination suppression threshold (measured in ms), is a measure of the degree of suppression the lead burst exerts on the lag burst in this ''precedence-effect'' situation. The present study found the following:

(1) For octave-band, random noise signals centered at 0.5, 2,0, or 3.0 kHz, the amount of discrimination suppression was determined primarily by the amount of spectral overlap between lead and lag sounds (experiment 1). As suggested by Blauert and Divenyi (1988), it appears that discrimination suppression (in the free-field) operates within frequency channels, and there is little or no remote spectral influence on discrimination performance.
(2) Localization strength for the three bands of noise employed in experiment 1 was independently measured (quantified as the MAA) and was found to decrease with increasing center frequency. The discrimination suppression results of experiment 1 were not consistent with Divenyi's (1992) ''localization strength hypothesis,'' which states that the amount of discrimination suppres-

sion should be directly related to the localization strength of the lead sound and inversely related to the localization strength of the lag sound.

(3) In order to test the localization strength hypothesis under conditions in which spectral (magnitude) features of the stimuli would have minimal effects, experiment 2 employed lead and lag stimuli that had the same center frequency (1.5 kHz) but differed in their rise times. The results of this experiment were consistent with Divenyi's hypothesis: the amount of discrimination suppression increased as lead stimulus rise time became more abrupt (i.e., as its localization strength increased) and decreased as lag stimulus rise time became more abrupt.

(4) Data from the present and previous experiments suggest that discrimination of lag-position information in a precedence effect situation is subject to at least two somewhat independent influences: (1) the degree of overlap of the magnitude spectra of the lead and lag sounds; and (2) the relative localization strengths of the two sounds. The relative weights of these two factors appear to depend on various stimulus factors that are only beginning to be understood. The available data are consistent with the following conclusions:

   (a) Localization strength is the predominant factor when spectrally nonoverlapping stimuli are presented over headphones and the cue to be detected is the ITD of the lag stimulus. Under these conditions low-frequency information is dominant over high-frequency information.

   (b) Spectral dominance is the predominant factor when spectrally nonoverlapping stimuli are presented in the free-field. In this case, a low-frequency lead stimulus does *not* suppress the discriminability of a high-frequency lag stimulus, presumably because ILD and spectral cues are available, which are *not* suppressed by the low-frequency lead. Suppression in this case operates within frequency channels and is governed by the amount of spectral overlap between lead and lag sounds.

   (c) When spectral overlap is held constant for stimuli presented in the free-field, abruptness of stimulus onsets governs the amount of suppression. This result suggests that localization strength may operate as a secondary factor for free-field stimuli and can be observed when the effects of spectral overlap are minimized.

(5) Other stimulus variables (such as bandwidth and duration of the signals) and task demands (such as whether the subject is tested on localization dominance or lag discrimination) may contribute to the relative weights of these two factors in precedence effect situations. Future study is required to fully describe the contributions of such variables.

ACKNOWLEDGMENTS

This research was supported in part by Grant No. DC00185. The authors are grateful to Dr. Raymond H. Dye (the associate editor), Dr. Pierre Divenyi, and an anonymous reviewer for their thoughtful input and constructive comments on an earlier version of this paper.

APPENDIX: MEASUREMENT OF MINIMUM AUDIBLE ANGLE FOR THE STIMULI EMPLOYED IN EXPERIMENTS 1 AND 2

The minimum audible angle (MAA) is a standard measure of auditory spatial resolution (e.g., Mills, 1958; Hartmann and Rakerd, 1989; Saberi *et al.*, 1991). It is typically measured by presenting two sounds in succession and determining the minimum angle between the two sounds for which the subject is just able to identify the directional change. Under the best circumstances (e.g., for a broadband stimulus presented from a region in front of the subject) the MAA in the horizontal plane is about 1° of arc (Mills, 1958).

In the present investigation MAAs were determined for the stimuli employed in experiments 1 and 2 in order to provide a measure of their relative ''localization strengths.'' Divenyi (1992) describes ''localization strength'' or ''localizability'' as the ''perceptual salience of spatial information''; he suggests that one way to quantify this salience (for headphone-presented dichotic stimuli) is to measure the interaural time difference (ITD) threshold. The smaller the ITD for the stimulus under investigation, the greater the subject's spatial resolution for that stimulus, and therefore the greater its localization strength is deemed to be. The equivalent measure for free-field stimuli is the MAA, and the same assumption applies: the smaller the MAA for a given stimulus, the greater its localization strength.

A. Method

The same five subjects were employed as in experiments 1 and 2, and testing was conducted in the same anechoic chamber. Stimuli were presented from a semicircular array of loudspeakers (JBL 8110H high-compliance full-range) positioned in front of the subject at ear level, 1.8-m distant. Angular separation between adjacent loudspeakers in the array was about 2° of arc. As in the main experiments, the room lights were extinguished during each threshold run.

In all cases, MAAs were determined at a reference azimuth of $+45°$, corresponding to the position of the lag source in the determination of discrimination suppression thresholds in the main experiments (Fig. 3). A two-down, one-up tracking procedure was employed. On each trial, two sounds were presented in sequence from two different loudspeakers on either side of the reference azimuth of $+45°$. The temporal interval between the two sounds was sufficient (300 ms) for the subject to hear both sounds distinctly. The subject's task was to state the order of the presentation of the two sounds (''left–right'' or ''right–left''), using a response box. Angular separation between the two loudspeakers was varied adaptively: two correct responses resulted in a decrease in the angular separation of the two loudspeakers on the next trial, and one incorrect response resulted in an increase in the angular separation on the next trial. The threshold (MAA) was estimated from the last eight of ten reversals of angular separation change. At least four thresholds were obtained in each condition for each subject.

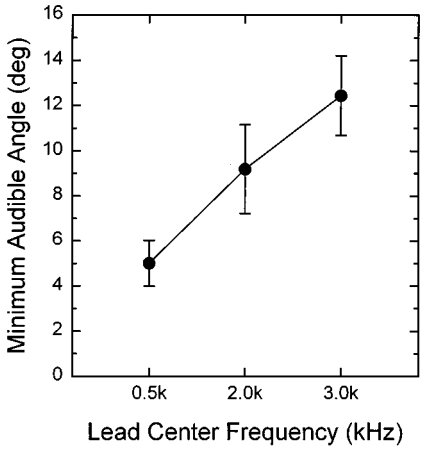


FIG. A1. Mean MAAs for five subjects for narrowband noise signals as a function of center frequency. These were the stimuli employed in experiment 1. Error bars show ±1 standard deviation about the means.

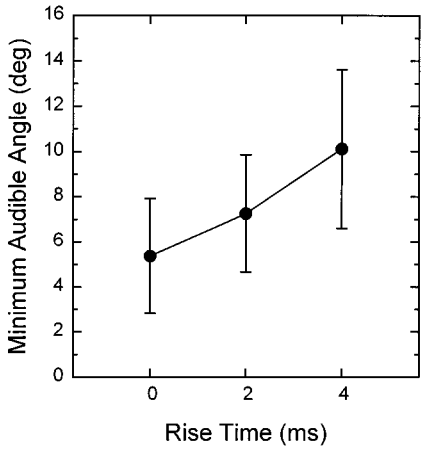


FIG. A2. Mean MAAs for five subjects for a 1.5-kHz tonal signal as a function of rise time. These were the stimuli employed in experiment 2. Error bars show ±1 standard deviation about the means.

B. MAA as a function of center frequency (experiment 1)

The stimuli in experiment 1 were 5-ms bursts of Gaussian noise, 1-octave wide, centered at 0.5, 2.0, or 3.0 kHz.[3] Mean MAA thresholds averaged across the five subjects are plotted in Fig. A1 as a function of center frequency. MAA threshold increased with increasing center frequency, a result that is in general agreement with Mills' (1958) findings with pure-tone stimuli over this frequency range. A one-way ANOVA indicated that the effect of center frequency on MAA threshold was significant [$F(2,8)=26.28$, $p<0.001$]. The 0.5-kHz noise had the lowest MAA threshold (5.0°), which indicates that it had the greatest localization strength of the three; the 3.0-kHz sound had the highest MAA threshold (12.4°), indicating that it was the least localizable of the three. These results are consistent with Divenyi's (1992) results obtained under headphones; using interaural temporal difference threshold as the measure of localization strength, he also found that the 0.5-kHz noise had the greatest localization strength while the 3.0-kHz noise had the least.

C. MAA as a function of stimulus rise time (experiment 2)

The stimuli in experiment 2 were 8-ms tone bursts (frequency: 1.5 kHz), whose rise times were 0, 2, or 4 ms. Average MAA thresholds across the five subjects for the three stimulus rise times are shown in Fig. A2. MAA threshold increased with increasing rise time: the more abrupt the onset of a pure tone, the more localizable it was, which is consistent with the results of Rakerd and Hartmann (1986). A one-way ANOVA indicated that the effect of threshold was significant [$F(2,8)=14.47$, $p<0.002$].

The implications of the changes in localization strength with center frequency (experiment 1) and rise time (experiment 2) for the interpretation of discrimination suppression are discussed in the main text.

[1]It should be noted that in this type of discrimination task, subjects may respond correctly by attending to *any* available cue; thus, it is not always possible to know whether they can actually detect position changes in the vicinity of the lag sound, or whether they respond to some other perceptual change in the overall stimulus (such as image width). Our subjects have reported that they do not notice changes in the position of the *lead* or composite sound in this task unless echo delay is very short (<1–2 ms); rather they attend to changes in the percept in the vicinity of the lag sound (Grantham, 1996; Yang and Grantham, 1997).

[2]The curious result in Divenyi's same-frequency condition (i.e., little or no suppression) may have been caused by the unusual stimulus configuration employed. In this case, the lag sound was an *inverted* copy of the lead sound; the correlation between lead and lag stimulus was thus −1.0.

[3]Filter settings for the 1-octave noises centered at 0.5, 2.0, and 3.0 kHz were 0.35–0.71, 1.4–2.8, and 2.1–4.2 kHz, respectively, using Kemo VBF 25.13 filters with nominal rolloffs of 90 dB/octave.

[4]In two conditions there was essentially no discrimination suppression, i.e., the thresholds were not measurable because the subjects could still perform the task at near 100% with echo delay less than 1.0 ms. In these cases, threshold was arbitrarily taken as 1 ms for further analysis.

Bernstein, L. R., and Trahiotis, C. (**1995**). ''Binaural interference effects measured with masking-level difference and with ITD- and IID-discrimination paradigms,'' J. Acoust. Soc. Am. **98**, 155–163.

Blauert, J. (**1983**). *Spatial Hearing: The Psychophysics of Human Sound Localization* translated by J. S. Allen (MIT, Cambridge, MA).

Blauert, J., and Divenyi, P. L. (**1988**). ''Spectral selectivity in binaural contralateral inhibition,'' Acustica **66**, 267–274.

Buell, T. N., and Hafter, E. R. (**1991**). ''Combination of binaural information across frequency bands,'' J. Acoust. Soc. Am. **90**, 1894–1900.

Clifton, R. K., and Freyman, R. L. (**1989**). ''Effect of click rate and delay on breakdown of the precedence effect,'' Percept. Psychophys. **46**, 139–145.

Clifton, R. K., Freyman, R. L., Litovsky, R. Y., and McCall, D. (**1994**). ''Listeners' expectations about echoes can raise or lower echo threshold,'' J. Acoust. Soc. Am. **95**, 1525–1533.

Divenyi, P. L. (**1992**). ''Binaural suppression of nonechoes,'' J. Acoust. Soc. Am. **91**, 1078–1084.

Divenyi, P. L., and Blauert, J. (**1987**). ''On creating a precedent for binaural patterns: When is an echo an echo?,'' in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 146–155.

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. (**1996**). ''Left-right asymmetry in the buildup of echo suppression in normal-hearing adults,'' J. Acoust. Soc. Am. **99**, 1118–1123.

Hafter, E. R., Saberi, K., Jensen, E. R., and Briolle, F. (**1992**). ''Localization in an echoic environment,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, K. Horner, and L. Demany (Pergamon, Oxford), pp. 555–561.

Hartmann, W. M., and Rakerd, B. (**1989**). ''On the minimum audible angle—A decision theory approach,'' J. Acoust. Soc. Am. **85**, 2031–2041.

Heller, L. M., and Trahiotis, C. (**1995**). ''Interference in detection of interaural delay in a sinusoidally amplitude-modulated tone produced by a second, spectrally remote sinusoidally amplitude-modulated tone,'' J. Acoust. Soc. Am. **97**, 1808–1816.
Heller, L. M., and Trahiotis, C. (**1996**). ''Extents of laterality and binaural interference effects,'' J. Acoust. Soc. Am. **99**, 3632–3637.
Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Litovsky, R. Y., and Macmillan, N. A. (**1994**). ''Sound localization precision under conditions of the precedence effect: Effects of azimuth and standard stimuli,'' J. Acoust. Soc. Am. **96**, 752–758.
McFadden, D., and Pasanen, E. G. (**1976**). ''Lateralization at high frequencies based on interaural time differences,'' J. Acoust. Soc. Am. **59**, 634–639.
Mills, A. W. (**1958**). ''On the minimum audible angle,'' J. Acoust. Soc. Am. **30**, 237–246.
Perrott, D. R., Marlborough, K., Merrill, P., and Strybel, T. Z. (**1989**). ''Minimum audible angle thresholds obtained under conditions in which the precedence effect is assumed to operate,'' J. Acoust. Soc. Am. **85**, 282–288.
Rakerd, B., and Hartmann, W. M. (**1986**). ''Localization of sound in rooms, III: Onset and duration effects,'' J. Acoust. Soc. Am. **80**, 1695–1706.
Saberi, K., and Perrott, D. R. (**1990**). ''Lateralization thresholds obtained under conditions in which the precedence effect is assumed to operate,'' J. Acoust. Soc. Am. **87**, 1732–1737.
Saberi, K., Dostal, L., Sadralodabai, T., and Perrott, D. R. (**1991**). ''Minimum audible angles for horizontal, vertical, and oblique orientations: Lateral and dorsal planes,'' Acustica **75**, 57–61.
Scharf, B. (**1974**). ''Localization of unlike tones from two loudspeakers,'' in *Sensation and Measurement: Papers in Honor of S. S. Stevens*, edited by H. R. Moskowitz, B. Scharf, and J. C. Stevens (Reidel, Dordrecht, The Netherlands), pp. 309–314.
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.
Stellmack, M. A. (**1994**). ''The reduction of binaural interference by the temporal nonoverlap of components,'' J. Acoust. Soc. Am. **96**, 1465–1470.
Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). ''The precedence effect in sound localization,'' Am. J. Psychol. **62**, 315–336.
Woods, W. S., and Colburn, H. S. (**1992**). ''Test of a model of auditory object formation using intensity and interaural time difference discrimination,'' J. Acoust. Soc. Am. **91**, 2894–2902.
Yang, X., and Grantham, D. W. (**1997**). ''Echo suppression and discrimination suppression aspects of the precedence effect,'' Percept. Psychophys. (in press).
Yost, W. A., and Soderquist, D. R. (**1984**). ''The precedence effect: Revisited,'' J. Acoust. Soc. Am. **76**, 1377–1383.
Yost, W. A., Wightman, F. L., and Green, D. M. (**1971**). ''Lateralization of filtered clicks,'' J. Acoust. Soc. Am. **50**, 1526–1531.
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), pp. 85–105.

Effects of three parameters on speaking fundamental frequency

Harry Hollien
Institute for Advanced Studies of the Communication Processes, University of Florida, Gainesville, Florida 32611

Patricia A. Hollien
Forensic Communication Associates, Gainesville, Florida 32601

Gea de Jong
Institute for Advanced Studies of the Communication Processes, University of Florida, Gainesville, Florida 32611

(Received 19 December 1996; accepted for publication 16 July 1997)

Speaking fundamental frequency levels and usage (SFF, $F0$) are of interest to many investigators who study human speech and voice. Substantial research in the area has been carried out; common foci include SFF as related to infant cry, age, gender, adolescent voice change, language, race, voice pathology, and so on. Yet there still are a number of relationships which are not well understood and three of them will be addressed in this project. They involve the long-held notions that (1) a secular trend exists with SFF being lowered over time, (2) the use of university students in research of this type will create bias because they are physically different from average individuals, and (3) SFF can vary systematically for different types of speech (especially for oral reading and extemporaneous speaking). Experiments assessing these questions were carried out, but only certain of the postulates were supported. That is, while some evidence of a secular trend was found, it appeared inconsequential during the past quarter of this century; second, although university students were found to be slightly larger than a cohort approaching the average population, only minor vocal differences were found. Finally, it was observed that, in general, oral reading resulted in higher mean SFF's than those for spontaneous speech. However, this difference was not robust and, due to reversals, the resulting metric did not appear to be of good predictive value for individual speakers.
 [S0001-4966(97)00911-9]

PACS numbers: 43.70.Gr [AL]

INTRODUCTION

Speaking fundamental frequency usage (SFF or $F0$) has been of particular interest to phoneticians and related professionals for many years. Indeed, fairly sophisticated investigations began some decades ago (Cowan, 1936; Lynch, 1934; Murray and Tiffin, 1934; Weaver, 1924); however, most of the relevant research has been carried out only since the middle of this century. Much has been accomplished but confusion yet exists about a substantial number of relationships. For example, a general listing of some of the more salient would include: (1) speaker characteristics of age, sex, health, race, education, language, temporal shifts, etc.; (2) the type of speech used (reading, spontaneous, extemporaneous, etc.); and (3) text materials (cold running speech, dramatic, informative, poetry, etc.).

As would be expected, it often is difficult to generalize from the data available because of the effects of the cited parameters, plus the sharply differing research approaches and measurement techniques employed. For example, the size of the populations studied has varied widely (from a half dozen individuals to several hundreds) and the age range within a group has extended from a few months to many decades. Sometimes median SFF is reported instead of the mean; standard deviation (variability) of the frequency distribution is usually reported (correctly) in tones or semitones, but sometimes it is found in Hertz. Finally, data extraction methods vary greatly; while adequate in most cases, these values occasionally are based on little more than estimates. Thus it would appear useful if some attempt were made at data coordination and if certain long-standing questions were addressed directly.

A. Areas exhibiting reasonable agreement

It would appear appropriate to review briefly those speaking fundamental frequency (SFF) relationships about which there appears to be some agreement. How gender and age affect SFF are areas where many of the major factors are understood and the data are relatively complete. For example, there is little doubt but that SFF levels are different for the sexes—at least, after puberty is reached (Fitch, 1990; Fitch and Holbrook, 1970; Higgins and Saxman, 1991; Krook, 1988; Künzel, 1989; Snidecor, 1994). The generalized curves found in Fig. 1 contrast these differences; they are based on the best available SFF evidence drawn primarily from data on Americans of European origin and Europeans where language appears not to be a biasing factor (see references listed below). Of course, graphed data such as these only provide a broad perspective of the $F0$ patterns. To illustrate, note the closely paralleling curves for infants and children; there appear not to be very great (gender) differences between them (if any at all)—at least before adolescent voice change takes place (Bennett, 1983; Eguchi and Hirsh, 1969; Fairbanks *et al.*, 1949a, 1994b; Hasek *et al.*, 1980;

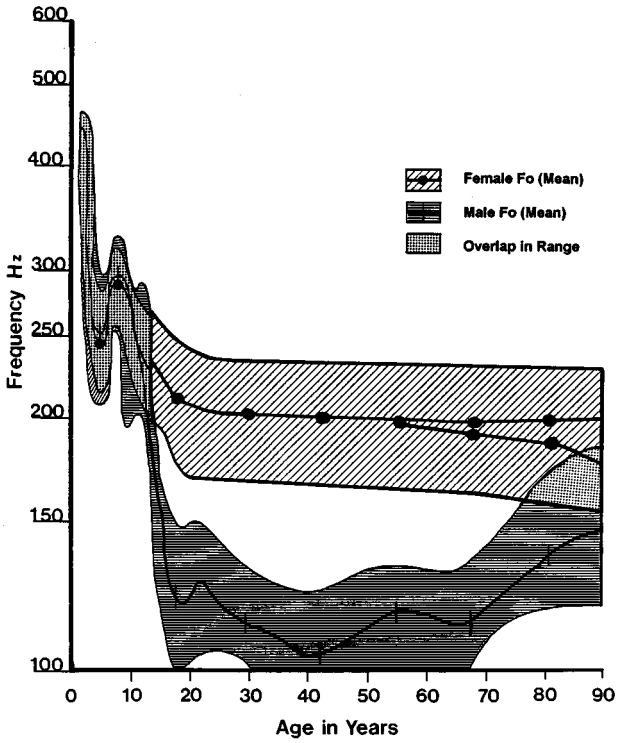


FIG. 1. Graphic portrayal of generalized data for male/female speaking fundamental frequency levels as a function of age. The data are plotted in Hertz; variation around the means is indicated by the hatched areas. Note that there now is a question as to whether female SFF remains constant or is lowered in older women.

McGlone and McGlone, 1972; Wheat and Hudson, 1988). It is during this latter period (i.e., puberty) where the sex-based differences in SFF level develop (Curry, 1940; Duffy, 1970; Hollien and Malcik, 1967; Hollien *et al.*, 1965, 1994).

Second, reconsideration of Fig. 1 will also reveal that there are enough $F0$-age data to provide a reasonably good perspective of how these factors *generally* covary over life (Brown *et al.*, 1991; de Pinto and Hollien, 1982; Fitch and Holbrook, 1970; Higgins and Saxman, 1991; Hollien, 1987, 1995; Hollien and Malcik, 1967; Hollien and Paul, 1969; Hollien and Shipp, 1972; Horii and Ryan, 1981; Krook, 1988; Linke, 1953/73; McGlone and Hollien, 1963; Mysak, 1959; Philhour, 1948; Pronovost, 1942; Ringel and Chodzko-Zajko, 1987; Saxman and Burk, 1967; Shipp *et al.*, 1992; Snidecor, 1951). However, many of the specifics about SFF and age are still lacking and some controversies are extant. For example, it is not yet known if $F0$ patterns for the elderly are random in nature (Chodzko-Zajko *et al.*, 1995; Ringel and Chodzko-Zajko, 1987) or follow orderly patterns such as those associated with the male–female coalescence theory (Hollien, 1987, 1995)—a model which suggests that older males and females tend to exhibit parameter shifts toward each other. The possibility that this theory may hold is reflected in the lower of the two $F0$ curves for older females (see again Fig. 1). It can be inferred by comparing the SFF levels reported by Honjo and Isshiki (1980) with those for younger Japanese women (Terasawa *et al.*, 1984; Tsuge *et al.*, 1987), and the suggestion of such a trend also is found in other studies (Krook, 1988; Mack, 1984; Stoicheff, 1981). On the other hand, the flat trend line for older women is based on the McGlone and Hollien (1963) data (especially when compared to Saxman and Burk, 1967), and the absolute values for older women reported by Krook and by Stoicheff. In any event, it was considered expedient to include both possibilities in the figure.

B. Areas exhibiting some confusion

Other SFF-linked relationships are even less well understood and the resulting confusion can impair comprehension of many specific SFF relationships. Moreover, they also tend to bias understanding in related areas. Examples include the SFF interface with race (Hollien and Malcik, 1967; Hudson and Holbrook, 1981; Mack, 1984; Morris, 1997; Wheat and Hudson, 1988; Yamazawa and Hollien, 1992), with language (Boë *et al.*, 1975; Hanley and Snidecor, 1967; Hanley *et al.*, 1966; Honjo and Isshiki, 1980; Majewski *et al.*, 1972; Terasawa *et al.*, 1984; Tsuge *et al.*, 1987; Yamazawa and Hollien, 1992), and with special behaviors such as smoking (Braun, 1994; Gilbert and Weismer, 1974; Sorenson and Horii, 1982; Stoicheff, 1981).

Even more important are the long-standing SFF controversies involving the effects of secular trends, subject class, and the types of speech used in research. If bias results from inadequate control of any of these factors, seemingly established relationships could be distorted.

(1) Secular trends. This term may be a misnomer; nevertheless, it has developed a certain currency. In this instance, it refers to a commonly held opinion that SFF will shift over time due to systematic changes in laryngeal structure or phonatory behavior. More specifically, it is argued that SFF for adults has been gradually lowered over recent years—and continues to drop. Yet, not very much data can be found to support this notion. The issue will be discussed in detail and, then, evaluated in the first section to follow.

(2) Subject class. A long-held opinion appears to be that measurement of certain vocal factors—in this instance SFF—will be biased by the type of subject studied. For example, SFF results are sometimes challenged because university students were used as subjects. It is contended that they do not adequately represent the overall population and, hence, they somehow bias the resulting data—in this case, that their SFF levels will be different from those found in the general population. While there are some data which suggest this postulate may be true, they are not very robust. Evaluation of this issue will be found in the second section of this report.

(3) Type of speech material. While it is without question that the use of certain types of speech material can produce SFF shifts (the contrasts among various emotionally laden samples constitute examples), the issue as to whether or not oral reading will result in different SFF levels than will extemporaneous speech has not yet been resolved. Again, data are sparse and variable. Hence, research attempting to resolve this issue was carried out (Sec. III).

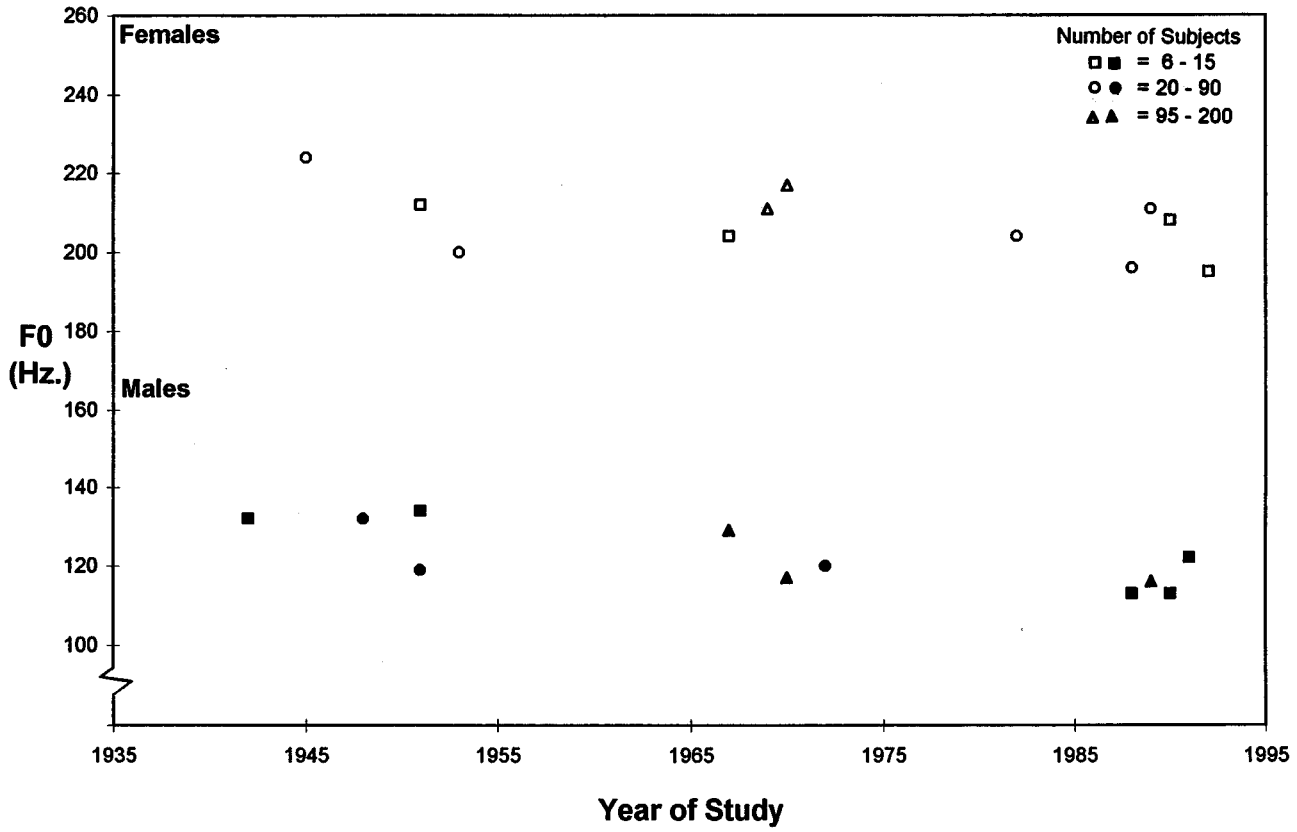


FIG. 2. Plot of mean speaking fundamental frequency data from 17 studies (American/European subjects only, drawn from de Pinto and Hollien, 1982; Fitch, 1990; Fitch and Holbrook, 1970; Hanley, 1951; Hanley and Snidecor, 1966; Higgins and Saxman, 1991; Hollien and Jackson, 1973; Hollien and Paul, 1969; Hollien and Shipp, 1972; Hollien *et al.*, 1982; Krook, 1988; Künzel, 1989; Linke, 1953/1973; Philhour, 1948; Pronovost, 1942; Snidecor, 1951; Yamazawa and Hollien, 1992). Note that group size is classified as ''small'' (squares), ''medium'' (circles), or ''large'' (triangles).

As may be seen, the ideas about these three factors may be due more to subjective thought than to the results provided by appropriate research. Nevertheless, it appears that they can be tested; they form the basis of the research to follow. Specifically, it is held that: (1) secular trends exist for speaking fundamental frequency usage; (2) the use of university students will bias SFF data as they can be expected to behave differently than normal healthy subjects; and (3) differences in SFF can result from the type of speech chosen for the research. Multiple experiments have been carried out to test these postulates; as stated, they will be considered serially and in the order cited above.

I. SECULAR TRENDS

The first issue involves the possibility that secular trends in speaking fundamental frequency level (SFF or *F*0) have occurred over the past half-century. Questions here are continually raised and available research can be used to support or contradict this position. For example, an SFF reduction of about 1.7 st. can be observed if Pronovost's 1942 data for young males are compared to those reported thirty years later by Hollien and Shipp (1972), i.e., 36.2 st. vs 34.5 st. Note also that his means (i.e., Pronovost's) are over 2.6 st. higher than are those reported by Krook in 1988 (36.2 st. vs 33.5 st.). A second example: a 4.7 st. difference separates the mean SFF data for 14-year-olds when Curry's 1940 value of 242 Hz (7) is compared to the 184-Hz level reported 25 years later by Hollien and Malcik (1967). However, when taken as a whole, the trends among the available data are not particularly compelling and the cited contrasts depend on selective comparisons. For example, little difference is found if Pronovost's 1942 data are compared to those reported in 1973 by Hollien and Jackson (i.e., 132 Hz vs 129 Hz) rather than Hollien and Shipp (1972). Nor can much difference be found if the mean data on middle aged women, reported in 1967 by Saxman and Burke (1967), are contrasted to those published 21 years later by Krook (1985) (i.e., 189 Hz vs 191 Hz).

In short, it is not at all clear if secular trends actually exist and (assuming they do) if they are robust. Indeed, if methodological differences are ignored, relationships such as those found in Fig. 2 can be plotted. In this case, they consist of those mean data, reported between 1940 and 1995, which we consider to be about as valid as any available—even when methodology and population size are taken into account. As may be seen, there is only a slight suggestion of overall SFF reduction for either the men or the women. Nevertheless, common opinion holds that a more robust SFF lowering trend exists. New data are provided in an effort to clarify the issue.

A. Data selection/generation

Several problems must be addressed if a reasonable set of comparisons are to be made over the 50-year period of interest. For example, the procedures which led to data col-

TABLE I. Summary table of the characteristics of the three groups of male subjects and the results of the SFF comparisons.

Parameter	Philhour 1948	Hollien/Shipp 1972	Present 1995
Number of subjects	24	25	25
Ages (years)			
Mean	YA[a]	24.4	23.4
Range	···	20–29	21–30
Mean SFF			
Hertz	132	120	114
Semitone level	36.2	34.5	33.6
Standard deviation (tones)	1.7	1.3	1.3

[a]YA=young adults.

TABLE II. Summary table of the characteristics of two groups of female subjects and the results of the SFF comparisons.

Parameter	Linke 1953	Present 1995
Number	27	18
Ages (years)		
Mean	YA[a]	22.5
Range	···	21–30
Mean SFF		
Hertz	200	198
Semitone level	43.3	43.2
Standard deviation (tones)	1.5	1.2

[a]YA=young adults.

lection should be both valid and compatible for those studies considered and the analysis approaches should be similar (see McKinney, 1965, and Hess, 1982, for general discussions about such equipment). Population size should be large enough to permit predictions and be of roughly the same type and size. Young healthy adults, 18–30 years of age attending a university were targeted as feasible groups, as were cohorts of 15–35 subjects. Further, it is known that the Phonellograph (see Cowan, 1936) and the Fundamental Frequency Indicator (Hollien and Tamburrino, 1965; Hollien, 1990) produce virtually identical data (Hollien *et al.*, 1994). Accordingly, studies where the investigator utilized one or the other of these devices were sought.

It was found that data from two early studies on males appeared compatible enough to permit reasonable contrast; they were Philhour (1948) and Hollien and Shipp (1972). Data were then collected in 1995 for a similar group of subjects in order to complete the triad (i.e., the current data also were obtained on 25 healthy university students). While it was a little more difficult to identify parallel materials for women, it was judged that the data collected by Linke in 1953 (but published in 1973) could be contrasted with a parallel group established for this purpose 42 years later. It is noteworthy also that all five groups were of reasonably similar size (range: 18–27 subjects), the speech materials were common to all and both equipment and measurement procedures were compatible.

B. Results

The basic data for the five groups cited may be found in Tables I and II. The males will be considered first. While only a slight lowering of SFF appeared to have occurred over the past 45–50 years, the shift proved significant at the 5% level ($F=10.87$; df=2,71, $F_{0.5}=3.98$). Possible variation within the trend slope was evaluated by application of a Tukey's post hoc test at the alpha=0.05 level. It was found that the Philhour data were significantly different from the others (required for significance >1.357; obtained=1.692), whereas the 1972–1995 contrasts were not (obtained=0.927). These relationships suggest that, if a secular trend does exist, it is confined primarily to the period prior to the 1970's.

Are there other factors—particularly external ones—which could explain the cited patterns? One source might be that of equipment fidelity. For example, early transducers and tape recorders were not of as high quality as those employed in the 1970s (and afterward); hence, low frequencies might have been damped in the Philhour research. On the other hand, the Phonellograph tends to detect more of the very low frequencies produced (vocal fry especially) than do processing systems such as FFI. Thus equipment limitations probably were not to blame as small restrictions in system frequency response undoubtedly were counterbalanced by those created by the processing equipment. About the only variable which could have *artificially* resulted in a downward secular shift is that of subject age. The Hollien–Shipp and present subjects (mean age: 24.4 and 23.4 years respectively) might be, on average, slightly older than Philhour's ''young adults'' and, hence, exhibit slightly lower SFF. Moreover, certain (additional) data are available which might be of relevance to this issue. They result from two studies that somewhat paralleled those reported but were not used as they did not meet the group size or speech sample criteria. The first was of a cohort of young males who served as controls for another project completed just after the Hollien and Shipp research (Hollien *et al.*, 1982). The mean age for this cohort of 41 subjects was 20.9 years and mean SFF was 125 Hz—a value more like Philhour's than the others. Further, the Hollien–Jackson (1973) SFF data were at a level nearly that of Philhour's and, with a mean of 20.3 years, that group was quite young also.

In short, it would appear that the data do not *fully* support the hypotheses that a secular trend exists; i.e., one where adult SFF levels have been systematically lowering over the past 50 years. Indeed, a reduction in SFF level of less than 0.06 st. per year over the period studied is simply too small a shift to support the notion of a robust secular trend.

The data for the female subjects (see again Table II) are not consistent at all with the hypothesis as stated. In this case, it may be seen that the two groups of young healthy university females, studied 42 years apart, exhibit almost identical SFF levels. There are so few differences (excepting time of data collection) extant between Linke's study and these new data, that this finding is rather compelling. Cer-

tainly there is no reason to reject a hypothesis that both groups are at the same level ($P = 0.73$).

C. Discussion

Certain relationships emerged when the data were further evaluated and contrasted with those reported in the early literature. First, it is now obvious that, if a powerful secular trend actually exists, much of it occurred prior to 1940 (witness Weaver, who in 1924 reported a mean $F0$ level of 312 Hz for females). Quite clearly, the restricted frequency response of the equipment used biased these very early studies. Yet a legitimate shift may have occurred, at least, when parallel data are considered. Second, another dividing point appears to have occurred, one which separates the two parts of the current half-century. Thus even if some sort of secular trend was operant during the first half of the period, there is little evidence that it still exists. The SFF differences sometimes reported during these more recent years appear related to (1) the types of subjects studied (for example, singers, nonsmokers), (2) problems with the subjects themselves (voice pathologies), and/or (3) behavioral states (subject stress, intoxication).

II. SUBJECT CLASS

The second issue involves the very common use of university students as subjects in phonetics research in general and studies of speaking fundamental frequency in particular. Concern is often expressed that individuals of this class exhibit characteristics different from those found in ordinary populations. The logic behind this opinion is that university students probably are larger, healthier, and smarter than individuals in otherwise similar cohorts. While this concern is rarely stated formally, it often surfaces at professional meetings, in reviews of proposals/publications and the like. Of course, it must be conceded that there may be a higher proportion of individuals who are ill or marginally competent among nonstudent populations and that students also may be more facile with tasks involving high levels of language skill and/or intellectual ability. On the other hand, these possible differences may not exist for other types of (behavioral) tasks; for example, those involving motor speech production and the rather specific characteristic of speaking fundamental frequency (SFF) usage.

A. Subjects

Speech materials for two young healthy male populations were gathered in the early 1970s. These studies were part of a much larger project designed to provide baseline speech/voice data (including SFF) for adults of all types. Some of the data from the first of these substudies have been published (Hollien and Jackson, 1973); however, no data from the other experiments have been reported as yet.

Subjects for the first study (i.e., Hollien and Jackson, 1973) were 157 male University of Florida students, aged 18–26 years (mean age 20.3 years), who were demonstrably healthy and exhibited no speech and hearing disorders. Nearly all of them had grown up in the Southeastern portion of the United States but none exhibited a marked regional dialect. The second cohort (unreported data) consisted of 142 healthy young men drawn from the enlisted ranks available at the Pensacola Naval Air Station. Their mean age and range was almost identical to the first group. All had been screened for speech, hearing, and voice disorders, and exhibited none. While, as with the university group, these subjects were laryngeally healthy, they included a slightly greater number of smokers than did the students. Those selected had completed high school but not attended college. Except for very few subjects in the Northeast group, none spoke with an observable regional dialect. These speakers were selected because they were thought to represent a normal, nonuniversity population—but a healthy one. While it is conceded that they probably were a little ''above average'' in some respects, their characteristics should be much closer to those for all young U.S. males than were those for the individuals in the university cohort. In short, the two groups studied were quite similar in most respects—that is, except for educational status. Note, also, that the second group was established to roughly parallel the demography of the United States. They were classified on the basis of the urban–rural population and by the four largest census regions; the relevant demographic data maybe found in Table III. As can be seen, these subjects exhibited about a normal urban–rural balance but were biased toward the Northeast, and away from the North Central region. However, it was judged that they reasonably represented the U.S. population and that they roughly matched the students except for the factor being tested. Finally, the two groups can be considered contemporary as the recordings for both were made over the same 14 month period.

TABLE III. Demographic data for the 142 subjects making up an average but healthy (military) group. Because these individuals were studied in the early 1970's, the demographic information is based on the 1960 U.S. census found in the 1968 Statistical Abstract. All values are in percent.

Area	U.S. population	Military subjects
(A) Region		
Northeast	24.9	36.6
North Central	30.7	19.0
South	28.8	21.8
West	15.6	14.4
Unclassified[a]	···	8.5
(B) Type		
Urban	69.9	68.4
Rural	30.1	31.6

[a]Primarily former military dependents.

B. Procedure

While speaking fundamental frequency was obtained for both oral reading and extemporaneous speaking, only the first of these two sets of data will be considered here. The material for the reading sample was a modified version of R.L. Stevenson's prose passage ''An Apology for Idlers.'' Subjects in both groups were provided ample practice time before recordings were made of this approximately 3-min passage. Height and weight were obtained just prior to the

TABLE IV. Basic data for the two relatively large populations of young adult males. The students were drawn from the University of Florida; the military cohort were young healthy adults who roughly parallel the population distribution of the United States.

Parameter	University	Military
Number	157	142
Age (years)		
Mean	20.3	20.5
Range	17.9–25.8	17.8–25.5
Height (inches)		
Mean	69.9	69.3[a]
Range	63.3–77.5	62.0–75.5
Weight (pounds)		
Mean	164.4	160.6[a]
Range	116.0–260.0	112.0–222.5
SFF (Hz)		
Mean	129.4	121.5
Range	92.6–178.1	95.0–159.4
Mean SFF (STL)	35.7	34.6
Standard deviation (T)		
Mean	1.6	1.6
Range	0.5–2.5	0.8–3.3

[a]Height and weight were not available for one subject.

speech trials. Recordings were made in an IAC sound-treated booth using a high quality Altec condenser microphone coupled to a calibrated Ampex 350 tape recorder. The resulting tapes were processed by IASCP's Fundamental Frequency Indicator (FFI-8) coupled to a dedicated computer (Hollien and Tamburrino, 1965; Hollien, 1990).

C. Results and discussion

Data for the two groups may be found in Table IV. As will be noted, they are almost identical in age and roughly similar in physical size, with the university students both slightly larger (on average by a half-inch and three pounds) and somewhat more variable relative to these factors. These data are consistent with the notion that university students will be larger than other healthy but otherwise average populations.

The data for speaking fundamental frequency level (see again Table IV) reveal over a semitone difference between the groups and this value is of statistical significance ($t=4.45$; $t_{0.05}=1.65$ for df>120). However, it can be said that, while the two groups appear different with respect to ($F0$) level, the shift is in a somewhat different direction than might be predicted on the basis of physical size. That is, while SFF does not appear to correlate very well with a subject's physical dimensions (Hollien *et al.*, 1994; Künzel, 1989; Majewski *et al.*, 1972), there are no data at all which suggest that smaller people will tend to exhibit lower $F0$ than will larger individuals. Thus even if a size–voice correlation could be argued, the trend here is in an unexpected direction—and, perhaps, in an inappropriate one. The only factor which might account for this shift could be a differential in the number of cigarette smokers in each group. While this variable was not controlled, the subjective impression was that there was only a slightly greater proportion of smokers in the military group and, in any case, subjects were young and could not have been long-term habitual smokers. Nevertheless, the possible effects of this factor cannot be ignored.

The values for the standard deviation of the distribution (s.d.) reflect the variability in $F0$ usage by the individual speakers. The virtually identical s.d. found for each of the two groups suggest that they are very much alike with respect to this parameter. Admittedly, the range is slightly lower and a little more restricted for the university students but this difference appears to be of little import.

III. ORAL READING VERSUS EXTEMPORANEOUS SPEAKING

Subjective impressions have led to the opinion that a person generally will exhibit a lower $F0$ when speaking extemporaneously than when engaged in the more formal activity of reading written material aloud. Sometimes results are questioned on the basis of which of these approaches was employed; in any event, the relationships cited here do not appear to be very well understood. Some research has been carried out (Fitch, 1990; Hanley, 1951; Hanley *et al.*, 1966; Higgins and Saxman, 1991; Hollien and Jackson, 1973; Snidecor, 1944), but no clear patterns have emerged. In some instances, the differences were quite small (Fitch, 1990; Hanley *et al.*, 1966; Morris, 1997); in others reversals were observed (Fitch, 1990; Hanley, 1951). There is also the possibility that different experimental procedures were employed even within a study. In one case, anyway, the subjects apparently sat when reading but stood when producing the extemporaneous speech. In short, it would appear reasonable to study the issue directly and possibly provide somewhat more definitive information about these relationships.

A. Procedure

Published data again were combined with new material to assess the issue of interest. However, all of the studies cited followed the same research protocols (i.e., those established by Hollien and Jackson, 1973). Not only were the same procedures used for both tasks but the (oral) reading passage was specifically balanced by an equally long extemporaneous speaking task (presentation order was counterbalanced). Four different experiments met these criteria; they were: (1) the 157 ''university'' males studied by Hollien and Jackson (1973); (2) the 142 ''military'' males cited above; (3) the 25 males from the secular trends study; and (4) 18 females also from that study. The contrasts were between ''The Apology for Idlers'' and a 2-min segment taken from 3 min of extemporaneous speech (responses to questions such as, ''What did you do on your last vacation?,'' ''What TV program is your favorite?'') in the first two investigations and between ''The Rainbow Passage'' and 40-s samples drawn from subjects' 2-min response to the same (neutral) questions in the second two. As stated above, subjects were young, healthy individuals between the ages of 18 and 30 years who stood or sat (depending upon which study) in fixed relationship to a microphone. All data reduction was accomplished by FFI processing.

TABLE V. Summary table contrasting speaking fundamental frequency measures for oral reading and extemporaneous speech.

Study	*N*	Reading	Extemporaneous	Diff. (ST)	Reversals (percent)
University males	157				21
Mean SFF (Hz)		129	123		
Mean SFF (STL)[a]		35.7	34.8	0.9	
s.d. (T)		1.6	1.6		
Military males	142				23
Mean SFF (Hz)		122	116		
Mean SFF (STL)		34.6	33.8	0.8	
s.d. (T)		1.6	1.7		
Current study: Men	25				28
Mean SFF (Hz)		114	110		
Mean SFF (STL)		33.6	33.0	0.6	
s.d. (T)		1.3	1.3		
Current study: Women	18				17
Mean SFF (Hz)		198	190		
Mean SFF (STL)		43.2	42.4	0.9	
s.d. (T)		1.2	1.1		
Means				0.8	22

[a]STL=semitone level; T=tones; ST=semitones.

B. Results and discussion

The contrasts between the speaking materials are best understood by consideration of Table V. First, it can be seen that the mean SFF level for the oral reading task was always higher than that for the more spontaneous speaking conditions. Moreover, the actual differences, while a little less than a semitone on average, are rather uniform—ranging only from 0.6 st. to 0.9 st. Thus it is tempting to conclude that the hypothesis, as stated, is supported by this research, especially since these differences were consistent for four separate studies, involving 341 subjects, carried out under reasonably precise conditions. This position is further strengthened by the results of an ANOVA ($F=74.74$, df $=1,3$, $F_{0.05}=10.13$). On the other hand, about one subject in five exhibited either no shift at all or a higher $F0$ for the extemporaneous speaking condition than for oral reading. This finding was not unexpected, especially since prior data on the issue were somewhat mixed. However, it precludes any strong predictions for *individual* subjects.

Additionally, it must be conceded that an uncontrolled variable could have biased the data for the spontaneous speech procedure; it involved the stress levels that might have been experienced by some subjects. That is, while the individuals studied responded to questions carefully structured for their neutrality, some of them appeared to become a little excited during their response (for example, they had ''absolutely loved'' the motion picture they were describing). Since it is reasonably well accepted that $F0$ can rise (for some subjects anyway) with increases in stress or certain types of emotions (Hollien, 1990), the possibility exists that such bias could have caused at least some of the reversals. In any event, it appears that higher group SFF can be expected for oral reading and there now is rather substantial support for this relationship.

IV. SUMMARY AND CONCLUSIONS

The three postulates evaluated were that: (1) a secular trend exists for SFF; one which is manifested as a lowering of talker $F0$ over time: (2) university students will exhibit somewhat different speaking behaviors than will average individuals; and (3) there are systematic differences in SFF level between reading aloud and speaking extemporaneously. Based on the obtained data, it must be concluded that none of them (except perhaps the third) can be supported at any but a minimal level. To be specific:

(1) While a statistically significant trend was noted for the SFF lowering of parallel groups over the past 50 years, there was no firm evidence that much (if any) of the shift occurred during the latter half of the period. Of course, the data presented were generated only on young adults. However, these findings were reasonably consistent with those from contemporary studies for individuals of other ages—especially when compensation was made for differences in population size/type, measurement procedure, data processing approaches and the like. If a physiologically based secular trend did occur it probably should be linked to periods prior to the 1970's. Moreover, it can be argued that recent data now are sufficiently extensive that direct comparisons with the earlier studies are no longer necessary (except for historical purposes).

(2) As expected, the university students studied were slightly larger than an otherwise comparable group of average young males. They also exhibited a slightly higher SFF. The problem here is that the vocal differences were not particularly compelling, appeared to be in an inappropriate direction and, although the SFF contrast between groups was statistically significant, both levels were well within the expected $F0$ range for young adults. Thus the second hypothesis hardly appears to warrant support. As a matter of fact, it can be argued that university students probably will make good subjects for studies of motor speech and voice even though their group $F0$ may be slightly higher than those for average subjects. They are less likely to exhibit minor deficits or pathologies and they certainly should be able to easily understand and carry out the tasks required of them.

(3) The third postulate was supported to some extent; however, with certain reservations. It appears that, under otherwise parallel conditions, most individuals will produce speech at a mean fundamental frequency which is higher for oral reading than for extemporaneous speech. This difference can be considered minimal for most applications and corrections should be easy to make (on a group basis, that is). It, also, must be stressed that, due to potential for reversals, these relationships do not result in good predictive metrics for individual speakers.

ACKNOWLEDGMENTS

The research upon which this report is based was supported by grants from the National Institutes of Health, the Dreyfus Foundation and the Office of Naval Research. The

authors also wish to thank Reva Schwartz, James Greene, Patricia Phillips, Allan Alderman and personnel at the U.S. Naval Air Station, Pensacola, Florida for their kind assistance and support.

Bennett, S. (**1983**). ''A 3-year longitudinal study of school-aged children's fundamental frequencies,'' J. Speech Hear. Res. **26**, 137–142.

Braun, A. (**1994**). ''The influence of cigarette smoking on vocal parameters,'' Proceed. ESCA Workshop on Speaker Recognition, Martiguy, 161–164.

Brown, W. S., Jr., Morris, R., Hollien, H., and Howell, E. (**1991**). ''Speaking fundamental frequency characteristics as a function of age and professional singing,'' J. Voice **5**, 310–315.

Boë, L. J., Contini, M., and Rabotofiringa, H. (**1975**). ''Statistique de la Fréquence Laryngienne,'' Phonetica **32**, 1–23.

Chodzko-Zajko, W., Ringel, R., and Vaca, V. (**1995**). ''Biogerontology Implications for the Communication Sciences,'' in *Communication in Later Life*, edited by R. A. Huntley and K. S. Helfer (Butterworth-Heinemann, Boston), pp. 3–21.

Cowan, M. (**1936**). ''Pitch and intensity characteristics of stage speech,'' Arch. Speech **1**, 7–85.

Curry, E. T. (**1940**). ''An objective study of the pitch characteristics of the adolescent male voice,'' Speech Monog. **7**, 48–62.

de Pinto, O., and Hollien, H. (**1982**). ''Speaking fundamental frequency characteristics of Australian women: Then and now,'' J. Phonetics **10**, 367–375.

Duffy, R. (**1970**). ''Fundamental frequency characteristics of adolescent females,'' Language Speech **13**, 14–24.

Eguchi, S., and Hirsh, I. (**1969**). ''Development of speech sounds in children,'' Acta Oto-Laryngol. Suppl. **257**, 1–51.

Fairbanks, G., Herbert, E., and Hammond, J. (**1949a**). ''An acoustical study of vocal pitch in seven and eight-year-old girls,'' Child Dev. **20**, 71–78.

Fairbanks, G., Wiley, J. M., and Lassman, F. (**1949b**). ''An acoustical study of vocal pitch in seven and eight-year-old boys,'' Child Dev. **20**, 63–69.

Fitch, J. L. (**1990**). ''Consistency of fundamental frequency and perturbation in repeated phonations of sustained vowels, reading and connected speech,'' J. Speech Hear. Dis **55**, 360–363.

Fitch, J., and Holbrook, A. (**1970**). ''Modal vocal fundamental frequency of young adults,'' Arch. Otolaryngol. **92**, 379–382.

Gilbert, H. R., and Weismer, G. G. (**1974**). ''The effects of smoking on the speaking fundamental frequency of adult women,'' J. Psycholinguist. Res. **3**, 225–231.

Hanley, T. D. (**1951**). ''An analysis of vocal frequency and duration characteristics of selected samples of speech from three American dialect regions,'' Speech Monog. **18**, 78–93.

Hanley, T. D., and Snidecor, J. C. (**1967**). ''Some acoustic similarities among languages,'' Phonetica **17**, 141–148.

Hanley, T. D., Snidecor, J. C., and Ringel, R. (**1966**). ''Some acoustic differences among languages,'' Phonetica **14**, 97–107.

Hasek, C. S., Singh, S., and Murry, T. (**1980**). ''Acoustic attributes of preadolescent voices,'' J. Acoust. Soc. Am. **68**, 1262–1265.

Hess, W. J. (**1982**). ''Algorithms and devices for pitch determination of speech signals,'' Phonetica **39**, 360–363.

Higgins, M. B., and Saxman, J. (**1991**). ''A comparison of selected phonatory behaviors of healthy aged and young adults,'' J. Speech Hear. Res. **34**, 1000–1010.

Hollien, H. (**1987**). ''Old voices: What do we really know about them?,'' J. Voice **1**, 87–101.

Hollien, H. (**1990**). *Acoustics of Crime* (Plenum, New York).

Hollien, H. (**1995**). ''The normal aging voice,'' in *Communication in Later Life*, edited by R. A. Huntley and K. S. Helfer (Butterworth-Heinemann, Boston), pp. 23–40.

Hollien, H., and Jackson, B. (**1973**). ''Normative data on the speaking frequency characteristics of young adult males,'' J. Phonetics **1**, 117–120.

Hollien, H., and Malcik, E. (**1967**). ''Evaluation of cross-sectional studies of adolescent voice change in males,'' Speech Monogr. **34**, 80–84.

Hollien, H., and Paul, P. (**1969**). ''A second evaluation of the speaking fundamental frequency characteristics of post-adolescent girls,'' Language Speech **12**, 119–124.

Hollien, H., and Shipp, T. (**1972**). ''Speaking fundamental frequency and chronologic age in males,'' J. Speech Hear. Res. **15**, 155–159.

Hollien, H., and Tamburrino, J. (**1965**). ''A fundamental frequency indicator,'' CSL Report 65-041 (Communication Sciences Laboratory, University of Florida, Gainesville).

Hollien, H., Green, R., and Massey, K. (**1994**). ''Longitudinal research on adolescent voice change in males,'' J. Acoust. Soc. Am. **96**, 2646–2654.

Hollien, H., Malcik, E., and Hollien, B. (**1965**). ''Adolescent voice change in southern white males,'' Speech Monogr. **32**, 87–90.

Hollien, H., Tolhurst, G. C., and McGlone, R. E. (**1982**). ''Speaking fundamental frequency as a function of vocal intensity,'' IASCP Report 027-82 submitted to the Dreyfuss Foundation (University of Florida, Gainesville).

Honjo, I., and Isshiki, N. (**1980**). ''Laryngoscope and voice characteristics of aged persons,'' Arch. Otolaryngol. **106**, 149–150.

Horii, Y., and Ryan, W. (**1981**). ''Fundamental frequency characteristics and perceived age of adult male speakers,'' Folia Phoniatr. **33**, 227–233.

Hudson, A., and Holbrook, A. (**1981**). ''A study of the reading fundamental vocal frequency of young black adults,'' J. Speech Hear. Res. **24**, 197–201.

Krook, M. I. P. (**1988**). ''Speaking fundamental frequency characteristics of normal Swedish subjects obtained by glottal frequency Analysis,'' Folia Phoniatr. **40**, 82–90.

Künzel, H. (**1989**). ''How well does average fundamental frequency correlate with speaker height and weight?,'' Phonetica **46**, 117–125.

Linke, E. (**1953**). ''A study of the pitch characteristics of female voices and their relationship to vocal effectiveness,'' Ph.D. dissertation, University of Iowa [later published in Folia Phoniatr. **25**, 173–185 (1973)].

Lynch, G. (**1934**). ''A phonophotographic study of trained and untrained voices,'' Arch. Speech **1**, 9–25.

Mack, L. (**1984**). ''A comparative analysis of linguistic stress patterns in Gullah (Sea Island Creole) and English speakers,'' unpublished MA thesis, University of Florida, Gainesville, FL.

Majewski, W., Hollien, H., and Zalewski, J. (**1972**). ''Speaking fundamental frequency characteristics of Polish adult males,'' Phonetica **25**, 119–125.

McGlone, R. E., and Hollien, H. (**1963**). ''Vocal pitch characteristics of aged women,'' J. Speech Hear. Res. **6**, 164–170.

McGlone, R. E., and McGlone, J. (**1972**). ''Speaking fundamental frequency of eight-year-old girls,'' Folia Phoniatr. **24**, 313–317.

McKinney, N. P. (**1965**) ''Laryngeal frequency analysis for linguistic research,'' Res. Rep. No. 14 (Communication Sciences Laboratory, University of Michigan, Ann Arbor).

Morris, R. J. (**1997**). ''Speaking fundamental frequency characteristics of 8- through 10-year-old White and African-American boys,'' J. Commun. Disord. **30**, 101–116.

Murray, E., and Tiffin, J. (**1934**). ''An analysis of some basic aspects of effective speech,'' Arch. Speech **1**, 61–83.

Mysak, E. (**1959**). ''Pitch and duration characteristics of older males,'' J. Speech Hear. Res. **2**, 46–54.

Philhour, C. W., Jr. (**1948**). ''An experimental study of the relationship between perception of vocal pitch in connected speech and certain measures of vocal frequency,'' unpublished Ph.D. Dissertation, University of Iowa.

Pronovost, W. (**1942**). ''An experimental study of methods for determining natural and habitual pitch,'' Speech Monogr. **9**, 111–123.

Ringel, R. L., and Chodzko-Zajko, W. (**1987**). ''Vocal indices of biological age,'' J. Voice **1**, 31–37.

Saxman, J. H., and Burk, K. W. (**1967**). ''Speaking fundamental frequency characteristics of middle-aged females,'' Folia Phoniatr. **19**, 167–172.

Shipp, T., Qi, Y., Huntley, R., and Hollien, H. (**1992**). ''Acoustic and temporal correlates of perceived age,'' J. Voice **6**, 211–216.

Snidecor, J. C. (**1944**). ''A comparative study of the pitch and duration characteristics of impromptu speaking and oral reading,'' Speech Monogr. **10**, 50–56.

Snidecor, J. (**1951**). ''The pitch and duration characteristics of superior females speakers during oral reading,'' J. Speech Hear. Dis. **16**, 44–52.

Sorenson, P., and Horii, Y. (**1982**). ''Cigarette smoking and voice fundamental frequency,'' J. Commun. Dis. **15**, 135–144.

Stoicheff, M. L. (**1981**). ''Speaking fundamental frequency characteristics of non-smoking female adults,'' J. Speech Hear. Res. **24**, 437–441.

Terasawa, R., Kakita, Y., and Hirano, M. (**1984**). ''Simultaneous measurements of mean air flow rate, fundamental frequency and voice intensity,'' Onseigngoigaku **25**, 189–207.

Tsuge, S., Kakami, K., and Fukaya, M. (**1987**). ''Speaker's height, weight and voice pitch for Japanese five vowels,'' Bull. Phonetic Soc. Jpn. **184**, 2–5.

Weaver, A. T. (**1924**). ''Experimental studies in vocal expression,'' J. Appl. Psychol. **8**, 23–56.

Wheat, M. C., and Hudson, A. I. (**1988**). ''Spontaneous speaking fundamental frequency of 6-year-old black children,'' J. Speech Hear. Res. **31**, 723–725.

Yamazawa, H., and Hollien, H. (**1992**). ''Speaking fundamental frequency patterns of Japanese women,'' Phonetica **49**, 128–140.

Simulating the effect of cochlear-implant electrode insertion depth on speech understanding

Michael F. Dorman[a)]
Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102 and University of Utah Health Sciences Center, Salt Lake City, Utah 84132

Philipos C. Loizou[b)]
Department of Applied Sciences, University of Arkansas at Little Rock, Little Rock, Arkansas 72204-1099

Dawne Rainey
Arizona State University, Tempe, Arizona 85287-0102

(Received 29 October 1996; revised 17 June 1997; accepted 1 July 1997)

Normally hearing listeners were presented with vowels, consonants, and sentences for identification through an acoustic simulation of a five-channel cochlear implant with electrodes separated by 4 mm (as in the Ineraid implant). The aim of the experiment was to simulate the effect of depth of electrode insertion on identification accuracy. Insertion depth was simulated by outputing sine waves from each channel of the processor at a frequency determined by the cochlear place of electrodes inserted 22–25 mm into the cochlea. The results indicate that simulated insertion depth had a significant effect on performance. Performance at 22- and 23-mm simulated insertion depths was always poorer than normal, and performance at 25-mm simulated insertion depth was, most generally, the same as normal. It is inferred from these results that, if insertion depth could be unconfounded from other coexisting factors in implant patients, then insertion depth would be found to affect speech identification performance significantly. [S0001-4966(97)02511-3]

PACS numbers: 43.71.Es, 43.71.Ky, 43.64.Me, 43.66.Ts [WS]

INTRODUCTION

Electrode arrays for cochlear implants are usually inserted 22–30 mm into the cochlea. The depth of insertion should make a difference in performance on tests of speech understanding. To see why, consider the case of an electrode array inserted approximately 27 mm into the cochlea. The most apical electrode will lie near the 350-Hz place and in, for example, an eight-channel prosthesis with logarithmic spacing of filters, the center of the first filter will be at 350 Hz. In this case, stimulation is delivered to the appropriate place in the cochlea. Consider, now, the case of an electrode array inserted only 22 mm into the cochlea. In this instance, the output of the first filter band is delivered to an electrode resting near the 800-Hz place in the cochlea. The mismatch between analysis frequency and stimulation frequency should make speech difficult to understand, since the phonetic segment characterized by energy at 350 Hz is different from the segment characterized by energy at 800 Hz.

Although we suppose that speech understanding will be best in a situation in which the stimulation from a filter band is directed to the ''correct'' place in the cochlea, clinical data do not provide overwhelming support for this position. Some patients with shallow insertions perform as well as patients with deep insertions (see, for example, Kileny *et al.*, 1992). Of course, in a clinical population the effect of electrode depth can not be assessed independently of coexisting conditions. For example, by chance a patient who has a deep insertion may not have stimulatable neural elements near the electrodes, while a patient with a short insertion may have neural elements near each electrode.

By testing normally hearing listeners it is possible to simulate the effects of insertion depth on speech understanding without encountering the confounds inherent in the use of implant patients as subjects. In the present experiment with normally hearing listeners, signals were first processed in the manner of a five-channel signal-processor for cochlear implants, for example, as in the Med El processor, and were then output as the sum of sine waves centered in the middle of each analysis band. Although no acoustic processing strategy can provide stimulation exactly like that experienced by an implant patient, because we cannot simulate the largely unknown dynamics of current spread in the cochlea, simulations like the one used here provide a starting point for trying to understand speech perception by means of cochlear implants (see, for example, Shannon *et al.*, 1995). For the baseline condition, the center frequency of the analysis band and the output signals for vowel, consonant, and sentence test materials matched exactly. In other conditions, the output signals were upshifted in frequency to simulate the effects of five electrodes spaced 4 mm apart inserted 22, 23, 24, and 25 mm into the cochlea. A 4-mm separation among electrodes was chosen because the implant patients with which we have the most experience use a 4-mm separation. To simulate a 22-mm insertion into the cochlea, the output of the first (418-Hz) filter band was delivered as an 831-Hz signal. The output of the second (748-Hz) filter was delivered at 1567 Hz. The outputs of the third (1.33 kHz), fourth (2.39 kHz), and

[a)]Electronic mail: mdorman@imap2.asu.edu
[b)]Electronic mail: loizou@ualr.edu

fifth (4.28 kHz) filters were delivered as sine waves at 2.84, 5.06, and 8.92 kHz, respectively. Greenwood's (1990) frequency-to-place equations were used to determine the sine wave output frequencies which simulated different electrode depths.[1] The aim of the experiment was to determine the simulated insertion depths, or differences between analysis frequency and output frequency, which would allow ''normal'' speech understanding for a five-channel prosthesis.

I. METHOD

A. Subjects

The subjects were female graduate and undergraduate students at Arizona State University who passed a hearing screening test at octave frequencies from 0.5–8 kHz. The subjects were paid for participation in the experiment. Nine subjects listened to the tests of sentence intelligibility. Five of the nine also listened to tests of consonant and vowel intelligibility. The subjects were included solely on the basis of availability for testing. All of the subjects had 12–15 h of experience listening to speech reconstructed as the sum of sine waves before participating in the experiment.

B. Stimuli

The test material included sentences, consonants in ''vCv'' environment, and vowels in ''hVd'' environment. The sentence material was from the H.I.N.T. test, presented without competing noise (Nilsson *et al.*, 1994). Examples of sentences in this test are, ''They met some friends at dinner,'' and ''Yesterday he lost his hat.'' A different set of ten sentences was used for each simulated insertion depth condition. The consonant test was the Iowa consonant test—16 consonants in /aCa/ environment spoken by a single male speaker (Tyler *et al.*, 1989). The stimuli were grouped into five blocks with each stimulus appearing once in a block. Stimulus order within each block was randomized. The vowel material consisted of the vowels in the words ''heed, hid, hayed, head, had, hod, hud, hood, hoed, who'd, heard.'' Each word was produced once by three men, three women, and three girls. The stimuli were drawn from a set used by Hillenbrand *et al.* (1995). The stimuli were presented in a completely randomized test sequence.

C. Signal processing

The sine-wave processor was implemented as follows. The signal was first preemphasized using a first-order FIR filter, and then bandpassed into five logarithmic frequency bands using sixth-order Butterworth filters. (A five-channel processor was simulated so that when the output frequencies were upshifted to simulate shallow insertions, the highest frequency would be under 10 kHz and would be within the frequency response of our Sennheiser HMD 410 headphones.) The center frequencies of the bands were at 418 Hz, 748 Hz, 1.33 kHz, 2.39 kHz, and 4.28 kHz. The envelope of the signal was extracted by full-wave rectification and low-pass filtering (second-order Butterworth) with a 400-Hz cut-off frequency. Sinusoids were generated with amplitudes equal to the rms energy of the envelopes (computed every 4 msec) and frequencies equal to the center frequencies of the bandpass filters. The sinusoids from each band were summed and presented to the listeners at a comfortable level.

TABLE I. Analysis filter center-frequency, output frequencies, and differences between analysis and output frequencies for simulations of electrode insertion depths of 22, 23, 24, and 25 mm. Entries are rounded to nearest whole number.

	Depth=22 mm			
	Analysis freq. ($f1$)	Output freq. ($f2$)	$f2-f1$	log ($f2/f1$)
channel 1	418	831	413	0.30
channel 2	748	1567	819	0.32
channel 3	1339	2844	1505	0.33
channel 4	2396	5065	2669	0.33
channel 5	4287	8924	4637	0.32
	Depth=23 mm			
	Analysis freq. ($f1$)	Output freq. ($f2$)	$f2-f1$	log ($f2/f1$)
channel 1	418	703	285	0.23
channel 2	748	1343	595	0.25
channel 3	1339	2456	1117	0.26
channel 4	2396	4390	1994	0.26
channel 5	4287	7751	3464	0.26
	Depth=24 mm			
	Analysis freq. ($f1$)	Output freq. ($f2$)	$f2-f1$	log ($f2/f1$)
channel 1	418	591	173	0.15
channel 2	748	1148	400	0.19
channel 3	1339	2118	779	0.20
channel 4	2396	3802	1406	0.20
channel 5	4287	6730	2443	0.20
	Depth=25 mm			
	Analysis freq. ($f1$)	Output freq. ($f2$)	$f2-f1$	log ($f2/f1$)
channel 1	418	493	75	0.07
channel 2	748	979	231	0.12
channel 3	1339	1823	484	0.13
channel 4	2396	3290	894	0.14
channel 5	4287	5840	1553	0.13

To simulate the 25-mm insertion depth, sine waves were output at 493, 978, 1823, 3290, and 5840 Hz. To simulate the 24-mm insertion depth, sine waves were output at 590, 1148, 2117, 3802, and 6729 Hz. To simulate the 23-mm insertion depth, sine waves were output at 703, 1343, 2456, 4390, and 7751 Hz. To simulate the 22-mm insertion depth, sine waves were output at 831, 1567, 2844, 5065, and 8924 Hz. The differences in analysis frequency and output frequency for each channel and for each simulated insertion depth are shown in Table I.

D. Procedure

The simulated insertion depth conditions were run in the order: Normal, 25, 24, 23, and 22 mm. This order was chosen to maximize the experience of the subjects before they were tested with the most difficult material. The tests were run in the order sentences, consonants, and vowels. This order was chosen to maximize the experience of the subjects before they were tested with the difficult task of multitalker vowel identification (see Dorman *et al.*, in press, for the results of other simulations with normally hearing subjects). Practice for the sentence material consisted of the presenta-

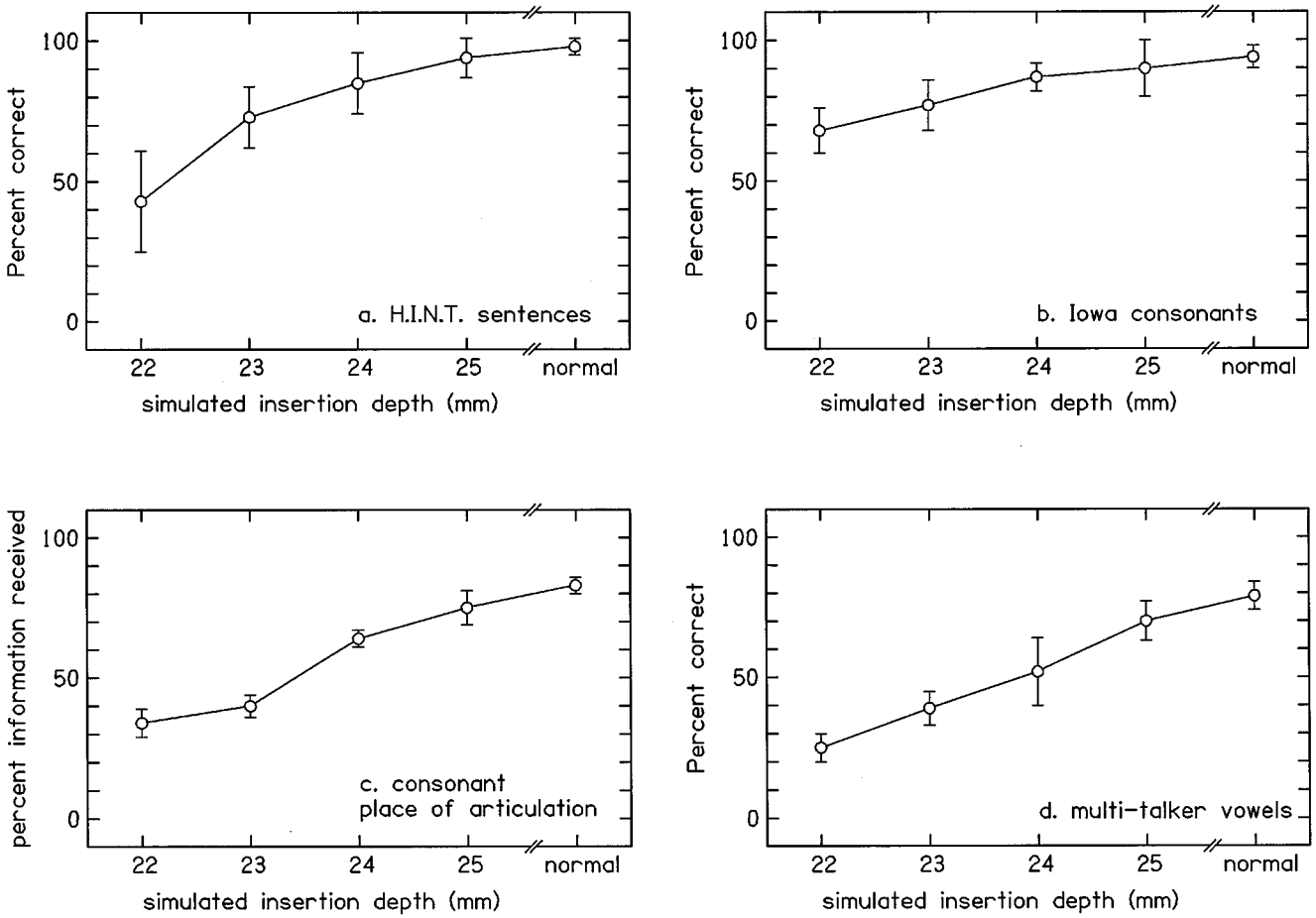


FIG. 1. Test scores as a function of simulated insertion depth. Error bars indicate ±1 standard deviation. ''Normal'' refers to a condition in which the center frequencies of the analysis bands and the output frequencies were identical.

tion of one list of ten sentences with concurrent visual display of the material for each simulation of insertion depth. Practice for the vowel and consonant tests consisted of two runs through the stimulus list with visual indication of item identity. In addition, the subjects were presented with one randomized set of stimuli with feedback of correct answers. Practice was given for each simulation of insertion depth.

All of the test materials were stored on computer disk and were output via custom software routines using MATLAB (Mathworks, Inc.) software and a 16-bit D/A converter. The signal level was approximately 72 dB *re:* peak vowel amplitude.

The subjects were tested in a sound attenuated room. Responses were collected with custom software using a computer display of response alternatives and a mouse as a response key. For the vowel and consonant material, the subjects were allowed to use a ''repeat'' key as many times as they wished. For the tests of word intelligibility in sentences, the subjects were presented a sentence once, and were instructed to repeat as many of the words as they could. Each word in the sentence was scored.

II. RESULTS

The results are shown in Fig. 1(a)–(d). A repeated measures analysis of variance for the sentence material indicated a significant effect for simulated insertion depth ($F[4,32] = 61.34$, $p < 0.0001$). *Post hoc* tests according to Scheffe (alpha = 0.05) indicated that the mean scores for the normal and 25-mm conditions (98% correct and 93% correct, respectively) did not differ. Both scores were significantly different from the 22-, 23-, and 24-mm conditions.

A repeated measures analysis of variance for the Iowa consonants indicated a significant effect for simulated insertion depth ($F[4,16] = 16.77$, $p < 0.0001$). *Post hoc* tests according to Scheffe (alpha = 0.05) indicated that the mean scores for the normal, 25-, and 24-mm conditions (94% correct, 90% correct, and 86% correct, respectively) did not differ. The scores were significantly different from the 22- and 23-mm conditions. A repeated measures analysis of variance for the feature ''place of articulation,'' scored for percent information received (Miller and Nicely, 1955), indicated a significant main effect for simulated insertion depth ($F[4,16] = 15.81$, $p < 0.0001$). *Post hoc* tests according to Scheffe (alpha = 0.05) indicated that the mean scores for the normal and 25-mm conditions (83% correct and 75% correct, respectively) did not differ. Both scores were significantly different from the 22-, 23-, and 24-mm conditions.

A repeated measures analysis of variance for the multitalker vowel material indicated a significant main effect for simulated insertion depth ($F[4,16] = 123.76$, $p < 0.0001$). *Post hoc* tests according to Scheffe (alpha = 0.05) indicated

that the mean scores for the normal condition differed from all other conditions.

III. DISCUSSION

There was a significant effect of simulated insertion depth for all test materials. We infer from this outcome that the depth of electrode insertion should make a difference in the speech perception abilities of cochlear implant patients. Performance in the 22- and 23-mm conditions differed significantly from the normal condition for all test materials. Thus relatively shallow insertions should result in relatively poor speech understanding when all other factors are held equal. For two of the three tests, including the test of sentence identification, performance in the 25-mm condition did not differ from that in the normal condition. This outcome can be rationalized by noting that the output frequencies in the 25-mm condition were within, or were just slightly beyond the edge, of the -3-dB point of the analysis filters. It may be the case that as long as the output frequencies are within the bandwidth of the analysis filters, differences between analysis filter center frequency and output frequency will not alter identification performance. At all events, insertions of 25 mm should be sufficient to allow the level of speech understanding which may be achieved with deeper insertions. Of course, this conclusion applies only to the electrode configuration we have simulated, i.e., the Ineraid electrode array with 5 electrodes at 4-mm intervals. Simulations of other electrode configurations may also be of interest.

In implant patients the effect of insertion depth may be realized in performance on two different measures. One is the time taken to reach asymptotic performance. The second is the level of asymptotic performance. Many implant patients report that for a period after the implant is activated, speech sounds ''high pitched'' or unnatural in a number of ways. Over time, patients usually report that speech sounds more ''normal'' and that speech is more intelligible, even though the parameters of stimulation have not changed. Given our results, it is possible that shallow insertion depths could be at least partially responsible for the reports of unnatural or ''high pitched'' speech at the time of processor activation and for poor initial intelligibility. (It is not likely that patients are sufficiently sophisticated to be able to tell the difference between an abnormally high pitch, *per se*, and upshifted formant frequencies.) It is important to note that if shallow insertion depth is responsible for the percepts described in the patient reports, then patients can learn to hear through, or compensate for, the distortion in frequency representation because, over time, most patients report that speech sounds more ''normal'' and is more intelligible.

Of course, insertion depth may also play a role in the level of terminal performance. Patients may be able to compensate for only distortions of a modest magnitude in the mapping of input frequency to cochlear place of stimulation. On this view, distortion in the mapping of input frequency to cochlear place of stimulation is one of many factors responsible for low levels of asymptotic performance among implant patients.

ACKNOWLEDGMENT

This research was supported by Grant No. RO1-000654-6 from the National Institute on Deafness and Other Communicative Disorders.

[1]We used a ''k'' of 1.0 in Greenwood's equation and assumed a membrane length of 35 mm. If a shorter membrane (e.g., 33 mm) is assumed then the perceptual consequences of short insertion depths would be reduced.

Dorman, M., Loizou, P., and Rainey, D. (**1997**). ''Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs,'' J. Acoust. Soc. Am. **102**, 2403–2411.

Greenwood, D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Kileny, P., Zimmerman-Phillips, S., Zwolan, T., and Kemink, J. (**1992**). ''Effects of channel number and place of stimulation on performance with the Cochlear Corporation multichannel implant,'' Am. J. Otolaryngol. **13**(2), 117–123.

Miller, G., and Nicely, P. (**1955**). ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–346.

Nilsson, M., Soli, S., and Sullivan, J. (**1994**). ''Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and noise,'' J. Acoust. Soc. Am. **95**(2), 1085–1099.

Shannon, R., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.

Tyler, R., Preece, J., and Lowder, M. (**1987**). ''The Iowa audiovisual speech perception laser videodisc,'' Laser Videodisc and Laboratory Report, Department of Otolaryngology, Head and Neck Surgery, University of Iowa Hospital and Clinics, Iowa City, IA.

The perceptual relevance of locus equations

David Fruchter and Harvey M. Sussman
Department of Linguistics, University of Texas, Austin, Texas 78712

(Received 3 July 1996; accepted for publication 22 July 1997)

Identification curves were estimated for the English consonants /b,d,g/ using five-formant CV synthetic stimuli comprehensively sampling the $F2$ onset–$F2$ vowel acoustic space in the vicinity of /b,d,g/ locus equations [H. Sussman *et al.*, J. Acoust. Soc. Am. **90**, 1309–1325 (1991)]. The stimuli included 10 English monophthongal vowel contexts, 11 levels of $F2$ onset per vowel, and 3 levels of $F3$ onset orthogonally varied with the $F2$ variables (10 vowels$\times$11 $F2$ onsets$\times$3 $F3$ onsets$=$330 stimuli). After brief training, each of six subjects, three male and three female, was presented eight trials of each of the stimuli, one or two trials per day over a period of several days. Systems of identification curves were visualized as identification surfaces situated in locus equation acoustic space and were overlaid with acoustic data from five male speakers in order to judge the degree of correspondence between perception and acoustic data. A chi square analysis was also performed in order to quantify the correspondence between the observed perception data and expected frequencies derived from the acoustic data. The results, when interpreted in terms of a dominance hierarchy hypothesis, strongly indicate $F2$ onset and $F2$ vowel, in combination, serve as important cues for stop consonant place of articulation. [S0001-4966(97)00111-2]

PACS numbers: 43.71.Es [WS]

INTRODUCTION

Locus equations are linear regressions of the frequency of the second formant transition sampled at its onset ($F2$ onset) on the frequency of the second formant sampled in the middle of the following vowel ($F2$ vowel) for a single consonant coarticulated with a range of vowels. The $F2$ onset is plotted on the y axis and the $F2$ vowel on the x axis. Locus equation variables, $F2$ onset and $F2$ vowel, are extremely highly correlated in natural speech (Sussman *et al.*, 1991, 1993, 1995). The phonetic origin(s) of this $F2$ transition endpoint correlation remain obscure, and theories concerning it are at this time speculative and controversial (Fowler, 1994; Sussman *et al.*, 1995, 1997). What is not controversial is the fact of the robust, stable correlation between locus equation variables across speakers and languages (Swedish: Lindblom, 1963; Krull, 1988, 1989; Canadian English: Nearey and Shammass, 1987; American English: Sussman *et al.*, 1991, 1995, 1997; Fowler, 1994; Thai, Cairene Arabic, Urdu: Sussman *et al.*, 1993; Spanish: Celdran and Villalba, 1995; Estonian: Eek and Meister, 1995). As the above references indicate, the high correlation between the locus equation variables has now been replicated in at least six different laboratories, and we may regard it as a well-established fact, and probably a language universal.

From their discovery by Lindblom during a study of vowel reduction in Swedish (Lindblom, 1963) up to the present day, locus equations have been employed descriptively, particularly as indices of degree of consonant–vowel (CV) coarticulation (Krull, 1989). In a departure from this descriptive tradition, Nearey and Shammass (1987), Sussman (1989) and Sussman *et al.* (1991, 1995) have postulated that the highly correlated locus equation variables, in combination, could be cues for consonant place of articulation perception. The basis for this concept is that the particular regression of $F2$ onset on $F2$ vowel depends on the consonant. Each regression function can be interpreted as defining an equivalence class of CV transitions, and each equivalence class maps to a particular consonant. The classes intersect to some extent, so they are only ''partially'' distinctive as to consonant; in those regions of intersection some other, nonlocus equation dimension (such as the burst, cf. Dorman *et al.*, 1977) is assumed to carry the distinctive information. Attributing to locus equations a role in consonant place of articulation perception offers a new perspective on a major portion of the noninvariance dilemma as it is manifested in speech perception (Sussman *et al.*, 1991).

Briefly, ''noninvariance'' means there is lacking a transparent map between a highly variable signal and a unit of the linguistic message. A particularly baffling, paradigmatic case of noninvariance has been the variability of $F2$ transitions associated with the stop consonants /b,d,g/ coarticulated in CV syllables with a range of vowels (Liberman *et al.*, 1954, 1967; Delattre *et al.*, 1955; Liberman and Mattingly, 1985). If a normalization routine were to be explicitly and mathematically derived for the variable $F2$ transitions within a place category, ''correcting'' for vowel-contexts, the function would do well to resemble a typical locus equation scatterplot. In the normal course of obstruent + vowel productions, the $F2$ onset–$F2$ vowel data self-organize into the equivalence classes mentioned above, which we have called ''vowel-normalized $F2$ transitions'' (Sussman *et al.*, 1995). We use this term in a descriptive, not in any formal or implemental sense.

A locus equation-based perspective on place of articulation perception stresses the organization of allophones of a phonetic category into a linearly correlated distribution of $F2$ coordinates in acoustic space. What makes this linear distribution of $F2$ coordinates nontrivial is that locus equa-

tions might not be articulatorily inevitable. Recent pilot work using the articulatory synthesis model, APEX (Stark *et al.*, 1996; Lindblom *et al.*, 1997), which generates an articulation-to-formant mapping, has shown that the full set of biomechanically possible articulatory configurations (e.g., lips, mandible, shape and position of tongue body and tip), for an apical closure for a given vowel context, is very broad. Thus, speakers' acoustic output ($\langle F2$ vowel, $F2$ onset$\rangle$) can vary, in principle, over an extensive acoustic range, and linearity across the vowel space is by no means assured. Linear relationships among $F2$ transition endpoints appear to be ''special'' in the sense of being implicitly ''selected'' by speakers from a large range of articulatory possibilities. Phonetically logical candidates to account for the linearity include (i) articulatory constraints operating to optimize speech production, and/or (ii) motor optimizations working in the service of perception. Part of the purpose of this study is to test the second possibility.

It should be noted that the locus equation theory of obstruent place of articulation perception across vowel contexts has been misinterpreted by some as implying that $F2$ onset in relation to $F2$ vowel is a *specifier* of consonant place of articulation rather than simply an important cue (Fowler, 1994). This is not the case. Rather, $F2$ onset in relation to $F2$ vowel is claimed to yield a partially distinctive cue (distinctive in many vowel contexts but not all) that typically functions as an important element of a polymorphic, redundant cue set. The locus equation-based hypothesis is not an invariance theory. It should not be so considered because it does not claim there is invariance in the signal. Rather, it claims there is modally a pattern of variation in the signal that can be readily mapped and usefully abstracted by a neural system. Rather than ''invariance,'' this notion of the relationship between waveform and phoneme might be termed ''mapable variance'' or ''orderly variance'' where ''mapable'' or ''orderly'' are to be defined strictly in terms of the feature-mapping properties of sensory neural systems (Sussman *et al.*, 1997).

In the following sections, some arguments for including locus equations in a theory of obstruent place of articulation perception across vowel contexts will be reviewed. This review will also motivate the research to be reported below, the thrust of which is to compare the acoustic production space of locus equations to the perceptual space of listeners instructed to identify stop consonants in synthetic CV continua orthogonally varying $F2$ onset and $F2$ vowel, the latter over a full range of vowel contexts. This experiment is designed to test the relevance of locus equation acoustic variables to the perception of stop consonants.

A. Rationale for the perceptual relevance of locus equations

1. Robustness of the phenomenon

One idea that locus equation data clearly refute is the Motor Theory notion that acoustic variability precludes establishing phonetic categories based on the acoustic signal:

> *Putting together all the generalizations about the multiplicity and variety of acoustic cues, we should conclude that there is simply no way to define a phonetic category in purely acoustic terms*. (Liberman and Mattingly, 1985, p. 12)
> *For if phonetic categories were acoustic patterns, and if, accordingly, phonetic perception were properly auditory, one should be able to describe quite straightforwardly the acoustic basis for the phonetic category and its associated percept. (ibid.)*

When viewed in terms of the linear and tightly clustered $F2$ onset–$F2$ vowel relationship, $F2$ transition variability by stop place category is about as orderly as one can find anywhere in Nature. As noted previously, the consonant-dependent $F2$ onset–$F2$ vowel relationships are stable and robust across speakers, dialects, languages and even in the face of articulatory perturbations produced by bite blocks (Sussman *et al.*, 1995). Discriminant analyses of speaker functions, using slope and y-intercept as predictor variables, have shown 100% correct classification of the functions into stop place categories [English: Sussman *et al.* (1991); Spanish: Celdran and Villalba (1995)]. Slope/y-intercept values may be somewhat modulated by speaker-induced changes in speaking style (Krull, 1989) or rate (Bakran and Mildner, 1995), but sufficient contrast among consonant place categories is maintained despite these modulations.

2. Perceptual studies of stop place of articulation cued by locus equation parameters

It has long been known that the $F2$ transition, as well as the noise burst, is an important cue for stop place (see Sussman *et al.*, 1991 for a complete review of this literature). Well before the locus equation metric was first derived (Lindblom, 1963), Liberman *et al.* (1954) manipulated $F2$ onset frequency, from a point 480 Hz below to 720 Hz above, the steady-state $F2$ resonances of seven vowel contexts in a two-formant synthesis study of stop consonant place perception. They concluded that ''*the results of this experiment show that the direction and degree of second-formant transitions can serve as cues for the aurally perceived distinctions among the stop consonants*'' (p. 9). While their Figure 5 showed respectable separability of ''b, d, g'' regions of perceptual space across vowel contexts, there were problematic contexts where the $F2$ transition was not very effective in cueing alveolars (e.g., /d, t/ perception in /i/ contexts).

In an effort to relate the results of this classic study to the locus equation parametrization of the $F2$ transition, Liberman *et al.*'s Figure 3 was replotted. We transformed their seven $F2$ vowel frequencies, their eleven $F2$ transition onsets (per vowel), and their identification values[1] to $\langle x,y,z\rangle$ coordinates to be compared with locus equation acoustic data obtained from five male speakers taken from Sussman *et al.* (1991). Figure 1 shows the obtained results. The identification data displayed as a gray scale are referred to as an ''identification surface.'' In analogy to a spectrogram, high rates of consonant identification are indicated by darker areas. The acoustic data to be compared are overlaid (as white circles). While two-formant pattern playback steady-state

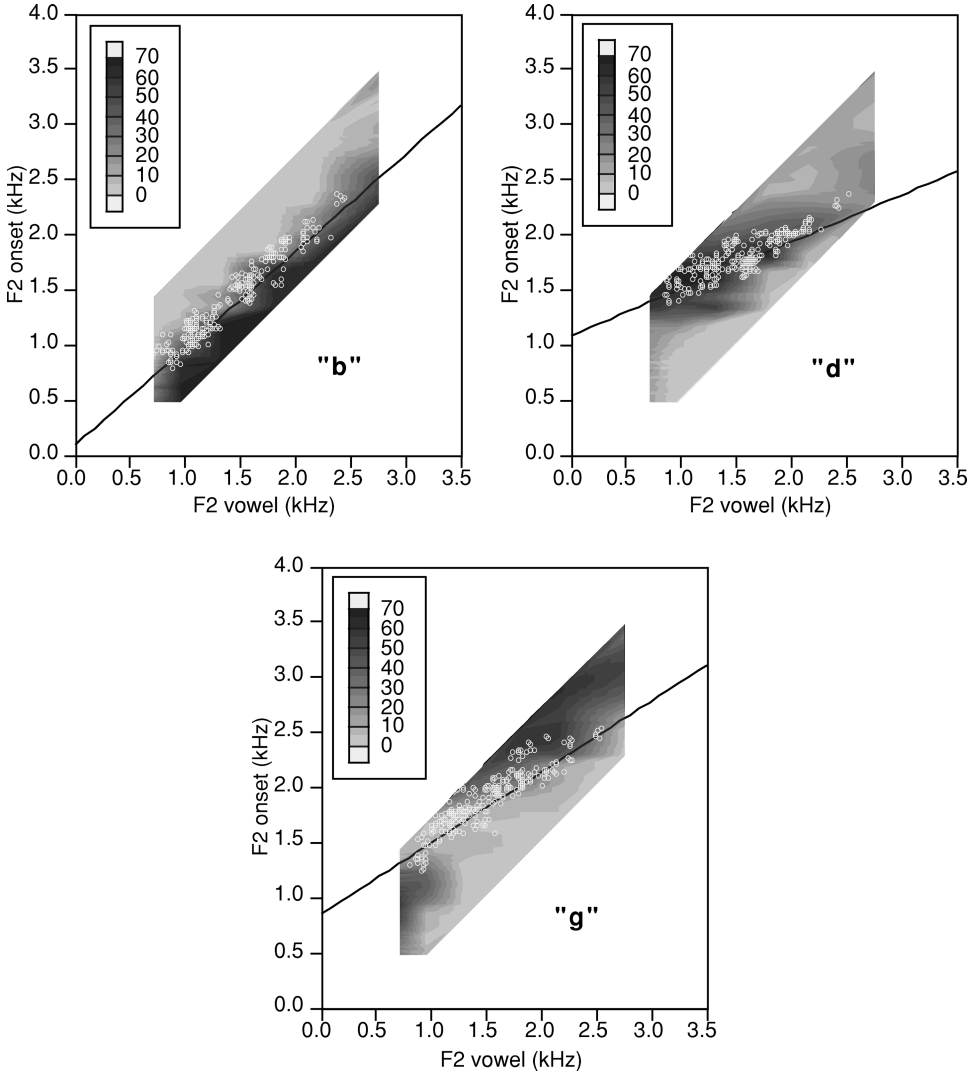


FIG. 1. Identification surfaces for ''b,d,g'' across seven vowel contexts. Data ($F2$ vowel, $F2$ onset, identification frequencies) transformed from Liberman *et al.* (1954) and displayed in locus equation coordinate space. Overlaid acoustic data from production (white circles) based on data from five male speakers taken from Sussman *et al.* (1991).

vowel target frequencies (essentially an $F2$-prime vowel) cannot be expected to exactly match $F2$ vowels from our speakers (Bladon and Lindblom, 1981), the correspondence between perception results and acoustic data are quite reasonable. The labial consonant identifications closely parallel the acoustic data. The ''d'' identifications also agree reasonably well with the superimposed locus equation data. The ''g'' results are noticeably weaker as areas of ''g'' perception predominate at $F2$ onsets $>$2250 Hz, and little acoustic $F2$ onset data falls in this region. Conversion of the Sussman *et al.* acoustic data to $F2$ prime values would increase the degree of correspondence to the Liberman *et al.* perceptual data in the upper-right quadrant of these figures, i.e., for /g/ in front vowel contexts.

A more recent perceptual test of locus equation parameters, $F2$ onset and $F2$ vowel, as cues for stop consonant place perception was performed by Eek and Meister (1995). Using Estonian CVV syllables beginning with voiceless unaspirated stops /p,t,k/ followed by nine long vowels, tokens from one speaker were used as stimuli for identification by 13 listeners. The $F2$ onsets ''without noise components of the transitional part of the burst'' were varied and presented to listeners for identification. In agreement with the Liberman *et al.* (1954) study, all rising $F2$ transitions were heard as ''p,'' while falling transitions were heard as ''t'' in back vowel contexts and ''k'' in front vowel contexts. Missing from the identification responses were alveolar ''t'' in front vowel contexts and velar ''k'' in back vowel contexts. The authors concluded that the best perceptual performance (75%–100%) resulted when the strongest peak of the burst was combined with a measure of $F2$ vowel-prime (rather than the $F2$ onset–$F2$ vowel parameters).

The research to be reported below was an attempt to replicate the results of Liberman *et al.* (1954), employing more realistic five-formant stimuli and an expanded vowel space. Also, advantage was taken of the presence of the $F3$ by independently manipulating it (three levels of $F3$ are employed—one level appropriate for [b], one for [d], and one for [g]). Correspondence of burstless CV stimulus identifications to acoustic locus equation data was evaluated to test the perceptual relevance of locus equation acoustic variables to stop place perception.

TABLE I. Values for ⟨$F2$ vowel, $F2$ onset⟩ associated with each of the points in Fig. 2. The arrangement of the table reflects that of Fig. 2—each column represents an $F2$ onset series for a given vowel context, and vowel contexts are arranged in order of ascending $F2$ vowel. The $F2$ onsets are stepwise equidistant on a Bark scale in the range from 480 Hz below to 720 Hz above the $F2$ vowel frequencies, except for the first level in the context [ɔ], which was constrained by the synthesizer's lower limit for $F2$.

Level	$F2$ onset (Y) values (Hz)									
11	1675	1788	1842	1905	1935	2345	2392	2511	2625	2900
10	1507	1620	1674	1737	1767	2181	2228	2349	2465	2745
9	1355	1467	1522	1584	1614	2029	2077	2199	2316	2599
8	1216	1328	1383	1445	1474	1889	1937	2059	2177	2462
7	1089	1200	1256	1317	1346	1759	1807	1930	2047	2332
6	970	1082	1138	1198	1228	1639	1687	1809	1926	2211
5	859	972	1028	1088	1117	1527	1575	1696	1813	2096
4	754	868	925	985	1014	1423	1470	1591	1707	1989
3	653	769	827	888	917	1325	1372	1492	1608	1887
2	565	674	733	795	824	1233	1280	1399	1514	1791
1	555	588	642	705	735	1145	1192	1311	1425	1700
	$F2$ vowel (X) values (Hz)									
	ɔ	o	a	ʌ	u	æ	ɛ	ɪ	e	i
	955	1068	1122	1185	1215	1625	1672	1791	1905	2180

I. METHODS

A. Synthetic speech stimuli

The stimuli were synthesized using the Klatt synthesizer, KLSYN88 (Klatt, 1980), running on a VAXStation under VMS. The parameter files that controlled the syntheses were derived from files generated by Klattalk, a text-to-speech program (Klatt, 1987). The process of creating the experimental stimuli began with ten Klattalk files, one for each bV syllable [bɔ],...,[bi]. The ten monophthongal vowel contexts, in order of ascending $F2$ vowel, were /ɔoaʌuæɛɪei/. The stimuli were given a male voice quality. The [b]'s were made burstless by zeroing the track for the variable parameter ''amplitude of burst.'' Several variables were held constant across all tokens: (i) the duration of all tokens was set to 300 ms, (ii) the pitch track of all tokens was set equal to that of the [bɔ] token, (iii) the durations of the $F2$ and $F3$ transitions were set to 48 ms for all tokens, and (iv) the VOT of all tokens was set to 6 ms. The $F2$ and $F3$ tracks from the end of their transitions (48 ms post-onset) through the end of the vowel (300 ms post-onset) were adjusted to reflect the mean $F2$ and $F3$ vowel values reported by Sussman *et al.* (1991) for a sample of ten adult males (Tables I and II). The adjustment was carried out by setting the end of the $F2$ and $F3$ transitions to the appropriate mean target value and then raising or lowering all following points of the $F2$ or $F3$ track to splice cleanly with the end of the transition. For example, the frequency of the $F2$ of [bi] 48 ms post-onset was changed from the original Klattalk value of 1994 Hz to the Sussman male mean value of 2180 Hz. Then, all subsequent values in the $F2$ track were raised 186 Hz (2180 minus 1994).

The $F2$ and $F3$ onsets were varied orthogonally. Three levels of $F3$ transition, one appropriate for [b], one for [d], and one for [g], were combined with 11 levels of $F2$ onset per vowel (11 $F2$ onset values per vowel×10 vowels×3 levels of $F3=330$ total stimuli for the experiment). The $F3$ onset values were adjusted to reflect the Sussman male means (Table II). As in the Liberman *et al.* (1954) experiment, the $F2$ onset series spanned a range from 480 Hz below to 720 Hz above the appropriate $F2$ vowel frequency. Intervals in the $F2$ onset series followed a stepwise equidistant plan such that there were, including the endpoints, 11 onset values per vowel separated by equal intervals on a Bark scale. Straight-line transitions were interpolated between $F2$/$F3$ onsets and their respective targets using the ''draw'' command of KLSYN88. The $F2$ vowel×$F2$ onset space sampled in this experiment is shown in Fig. 2. Table I provides frequency values for the stimulus coordinates shown in Fig. 2.

The waveforms based on the completed. DOC files were

TABLE II. Values for $F3$ vowel and for $F3$ onset, by vowel, for the three levels of $F3$ in the experiment, b-like $F3$, d-like $F3$, and g-like $F3$.

	$F3$ onset values (Hz)									
b-like	2381	2313	2403	2420	2270	2343	2365	2400	2357	2460
d-like	2418	2406	2458	2527	2359	2566	2559	2569	2557	2667
g-like	2170	2178	2195	2227	2172	2551	2617	2664	2632	2904
	$F3$ vowel values (Hz)									
	ɔ	o	a	ʌ	u	æ	ɛ	ɪ	e	i
	2471	2357	2518	2546	2284	2460	2483	2473	2518	2728

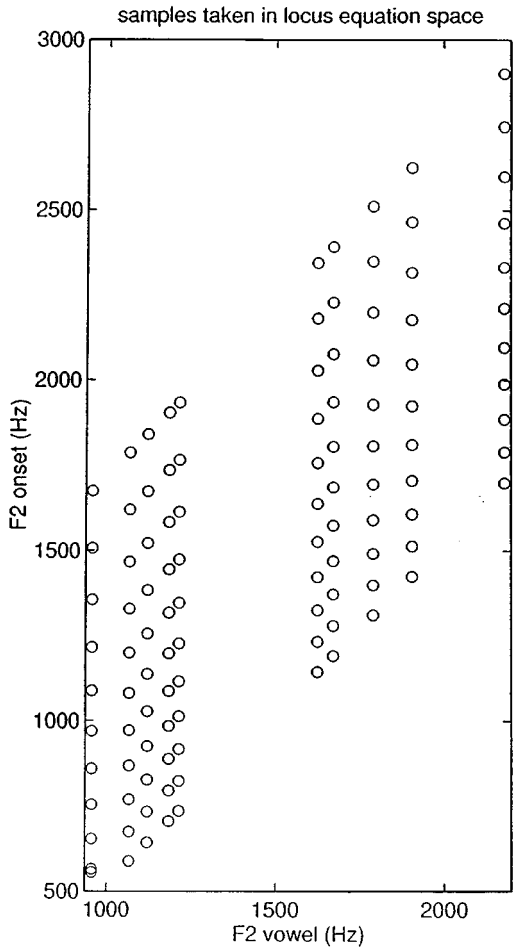


FIG. 2. Distribution in $F2$ locus equation acoustic space of the synthetic CV stimuli for the identification experiment.

synthesized, and the peak amplitude of each token was set to a standard by multiplying the waveform by an appropriate factor. Stimuli were recorded from digital files onto reel-to-reel tape (Tandberg TD 20A tape deck). Two tapes were made, each consisting of three blocks of 110 stimuli. Each block comprised a single level of $F3$ with 10 vowel contexts$\times$11 levels of $F2$ onset. Order of stimuli within a block was randomized. Different within- and between-block orders were employed on the two tapes.

B. Consonant identifications

Stimuli were delivered to subjects via a reel-to-reel tape deck (Tandberg TD 20 A) connected to an amplifier and headphones (Realistic stereo mixing console, Adcom GFA-535L amplifier, splitter box, beyerdynamic DT100 headphones) in a quiet room. The six subjects were all graduate students between the ages of 25 and 40 years old. All subjects reported normal hearing. All were native speakers of an American or Canadian English dialect. Subjects 1, 3, and 4 were male, while subjects 2, 5, and 6 were female. With the cooperation of the experimenter, subjects were asked to set their own level to be comfortable but moderately loud. An oscilloscope reading for a calibration tone played at this level setting was recorded, so the chosen level could be duplicated for subsequent sessions with a given subject. Subjects were trained in the task sufficiently (220 randomly selected stimuli) to become comfortable with a relatively rapid stimulus presentation rate, adopted to prevent subjects from becoming bored later with the task. No feedback was given during training except to initially assure subjects that by the end of training they would be adapted to the task rate. Initially, the trials seemed speeded to some subjects, but by the end of training all acknowledged the presentations were comfortably paced. In order to limit fatigue, a maximum of two tapes (720 stimuli) were tested on a given day. Thus, each subject listened to eight replications per stimulus over a minimum of four days. Although fewer subjects were used than in Liberman *et al.* (1954), each subject was run through more trials to discern whether there was significant between-subject variation. Subjects were asked to identify each stimulus as most similar to ''b,'' ''d,'' ''g,'' ''w,'' or ''no consonant.'' The responses ''w'' and ''no consonant'' were not part of the response repertoire of Liberman *et al.* (1954), but were added to this replication design, after the first author had listened to the stimuli, in order not to force subjects to make sometimes nonsensical identifications. Subjects were warned to use the last alternative only if there was really no consonant, not if the consonant was simply ambiguous.

C. Analysis

Identification frequencies were tabulated and entered along with $F2$ onset and $F2$ vowel of the eliciting stimuli into graphics software (MATLAB version 4.2c.1) to produce the ''identification surfaces.'' Each identification surface is based on 110 triplets of the form $\langle F2$ vowel, $F2$ onset, identification frequency$\rangle$—the identification frequencies are added on as the z values. The software interpolates between sample points in a manner analogous to dropping a cargo net over a 3-D scatterplot of the identification data. The method of shading, known as Gouraud shading, is piecewise bilinear; the shading of each patch varies linearly and interpolates the corner values of the rhomboids (The MathWorks, Inc., 1992).

A chi square method used to statistically compare the identification results with the acoustic data is described in Sec. II C.

II. RESULTS

The results are first presented as identification surfaces along with overlaid locus equation acoustic data. It is with this type of display that the correspondence of perceptual responses to locus equation acoustic data can best be judged. A chi square analysis of the degree of correspondence between identification and acoustic data appears in a later section. There is also a brief presentation of the effects of $F3$.

A. Identification surfaces

Recall from Fig. 2 that there were 110 combinations of $F2$ vowel and $F2$ onset composing a set of stimuli for a given level of $F3$. The format of Fig. 2 is that of a normal locus equation acoustic space, so that each vertical line of points is an $F2$ onset series (11 levels) for a given vowel context, and the ten vowel contexts are arranged horizontally

in order of ascending $F2$ vowel frequency. Each of the points in Fig. 2 represents a stimulus in the perception experiment, so that there is identification data for each of them. If the identification data are graphed on a z axis, we can create a surface by interpolating between adjacent triplets $\langle x,y,z\rangle$. The overall effect is as if the individual $F2$ onset series identification curves for a given consonant identification were slices of bread, each slice being a different vowel context, and the slices were assembled into a loaf, which we then view from above. These surfaces have the decided advantage of being able to simultaneously display all vowel contexts for a given consonant identification and level of $F3$. For translation to a two-dimensional medium, the z axis is coded via a gray scale—the higher the rate of identification, the darker the shading—similar to the intensity coding of a sound spectrogram. All figures show subject-pooled data (i.e., z ranges from 0–48).

Although $F3$ was a variable in this experiment, it had little effect, so only surfaces for appropriate levels of $F3$ (b-like for ''b,'' d-like for ''d,'' or g-like for ''g'') will be shown, as they are representative. (The effects of $F3$ did show an interesting overall pattern, which will be described after the main result.) For the consonant identifications ''b,'' ''d,'' and ''g,'' corresponding token-level locus equation acoustic data from five male native speakers of English producing ''beat, bit, bait,... , deet, dit, date,... , geet, git, gate,...'' five times each (data taken from Sussman *et al.*, 1991) are overlaid (white circles). The overlays allow appraisal of the correspondences between the distribution of the acoustic data and features of the perception data.

There are clear peaks and valleys in the identification surfaces, a peak (darker region) representing a high rate of identification as a given consonant. Turning first to the ''b'' surface (Fig. 3), it can be summarized by saying that a high rate of ''b'' response is indicated when the $F2$ onset is lower than the $F2$ vowel. In the back vowel region ($F2$ vowel ca. 1000 Hz), a high rate of ''b'' identification is achieved even when the $F2$ onset is slightly above the $F2$ vowel. The overall effect for ''b'' identifications is of a massif, or ridge, rather than of a single peak. The overlaid acoustic data correspond very well with this ridge of ''b'' perception.

The ''d'' surface is shown in Fig. 4. The ''d'' peak in the back vowel region ($F2$ vowel$<$c. 1300 Hz) exactly complements the ridge for ''b'' i.e., it is clear that the ''b'' and ''d'' curves do not overlap in back vowel contexts. On the other hand, in the front vowel contexts ($F2$ vowel$>$c. 1600 Hz) the ''d'' peaks are lower than in the back vowel region and overlap the ''b'' ridge. Noting the attenuated intensity of the ''d'' peaks in the front vowel region, there is nevertheless a good correspondence between the ''d'' identification peaks and the overlaid [d] acoustic data.

Referring to the ''g'' surface (Fig. 5), in front vowel contexts the peak region for ''g'' complements the peaks for ''b'' and ''d.'' Actually, it appears there may be, near the joint boundaries, cases of ''triple-point'' front vowel context stimuli that are almost perfectly confusable among the three stops. The front vowel peak for ''g'' is attenuated above an $F2$ vowel of c. 1750 Hz. The back vowel region is characterized by a relative lack of ''g'' responses, except for one

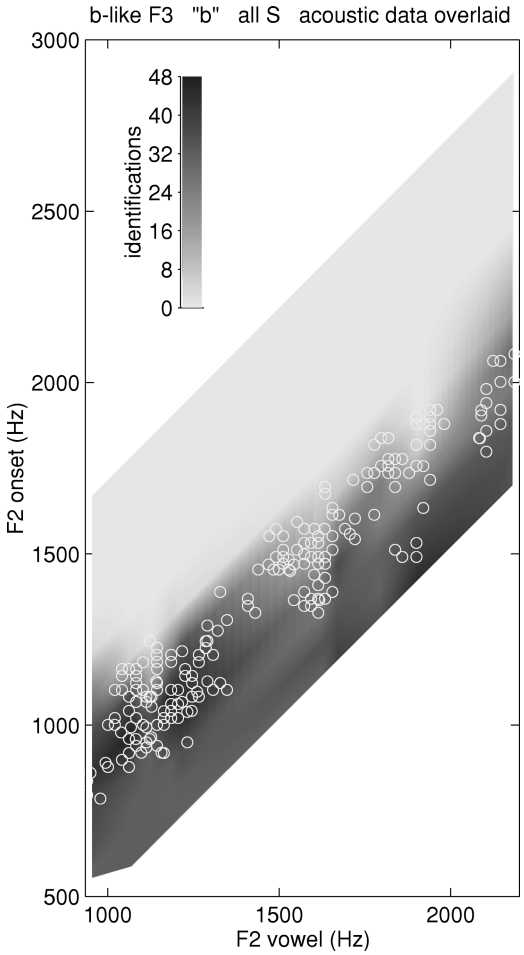


FIG. 3. Identification surface for ''b'' overlaid with acoustic data (white circles) from production of bVt syllables [taken from Sussman *et al.* (1991)].

vowel context (/a/, in the region of $F2$ vowel at c. 1100 Hz). The correspondence between the ''g'' peak and the overlaid acoustic data is fair in the front vowel region. The back vowel correspondence is problematic, inasmuch as there is very little ''g'' identification but plenty of [g] acoustic data distributed in the region.

The patterns of ''w'' and ''no consonant'' responses were not sufficiently interesting to warrant a graphical presentation. In general the ''w'' surfaces were relatively featureless. This turned out to be a case of between-subject variation, with two subjects showing great activity against a background of four showing none. The few ''w'' responses were concentrated in the back vowel region at the lower limit of the $F2$ onset series, where they overlap the ''b'' curve in subject-pooled data, or cross over the ''b'' curve in those subjects that gave a ''w'' response. The ''no consonant'' surfaces also showed little activity. Subjects were instructed to use this response sparingly. There seemed to be a gradient from the back vowel region to the front vowel region, with higher and more widespread activity with increasing $F2$ vowel. As might be expected, ''no consonant'' activity at the lower $F2$ vowel values (below ca. 1750 Hz) was concentrated among those stimuli having a relatively flat $F2$ consonant-vowel transition. Above ca. 1750 Hz, the CV transition trajectory seemed to have less influence; it seemed

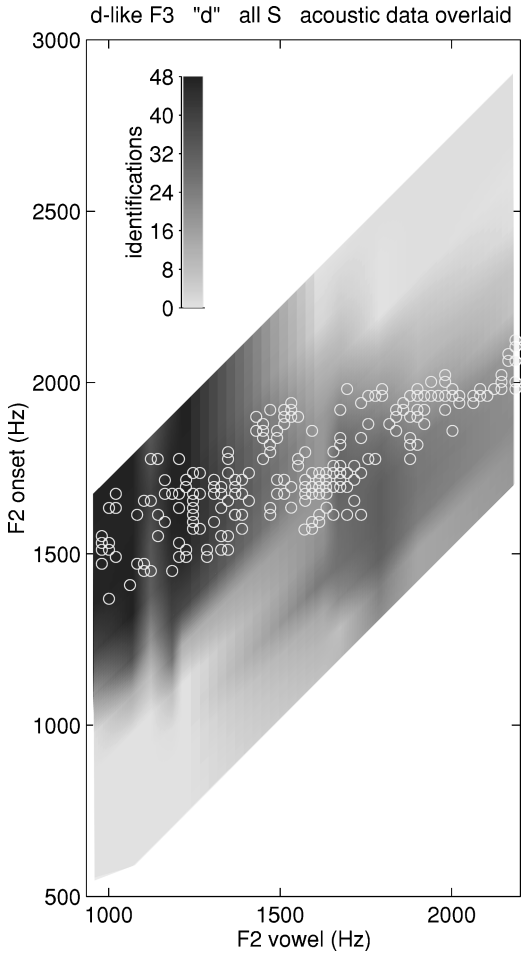


FIG. 4. Identification surface for ''d'' overlaid with acoustic data (white circles) from production of dVt syllables [taken from Sussman *et al.* (1991)].

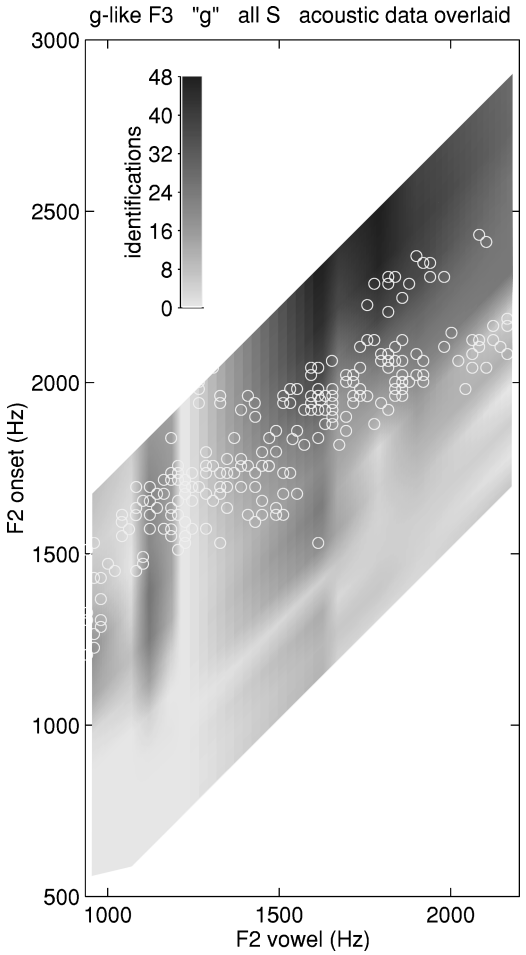


FIG. 5. Identification surface for ''g'' overlaid with acoustic data (white circles) from production of gVt syllables [taken from Sussman *et al.* (1991)].

especially difficult to evoke a stop identification with the vowel context /i/, even for a wide range of $F2$ transitions. As with ''w,'' there was considerable between-subject variation in the frequency and distribution of the ''no consonant'' responses, with two subjects particularly prone to use it compared to the other four.

B. Effects of *F3*

Table III presents a summary of the effects of level of $F3$ on stop place perception. A certain degree of symmetry, of an antiparallel type, is evident in the table. Particularly, there was no effect of the $F3$ condition on ''b'' identification in back vowel contexts and on ''g'' identification in front vowel contexts, while there were effects on ''d'' versus ''g'' identification in back vowel contexts and on ''b'' versus ''d'' identifications in front vowel contexts. In comparing these patterns of $F3$ effects with locus equation acoustic data, a strong relationship emerges, namely, that there is a lack of $F3$ effects in those regions in which there is a lack of overlap between the different stop places of articulation (back vowel [b] and front vowel [g], while there are tradeoff effects between the overlapping stops in the region of their overlap ([d] and [g] overlap in back vowel space, [b] and [d] in front vowel space). These tradeoffs are in the natural directions, with g-like $F3$ elevating ''g'' versus ''d'' identifications, and b-like $F3$ elevating ''b'' versus ''d'' identifications.

C. Quantitative comparison of acoustic and identification data

In an attempt to quantify the degree of relationship, or in some sectors of locus equation space, the lack of direct relationship, between locus equation acoustic data and the identification frequencies, a chi square analysis was performed.

Expected frequencies were generated by a model in which identification frequencies directly reflect the distribution of acoustic data local to a stimulus in the identification experiment. Locality was defined in terms of nearest neighbors in locus equation acoustic space. For each $\langle F2$ vowel, $F2$ onset$\rangle$ point sampled in the identification experiment

TABLE III. Summary of the effects of level of $F3$ on the rate of consonant identifications, by frontness-backness of the vowel context.

Effect on	For back V contexts	For front V contexts
''b''	none	b-like elevates, g-like depresses
''d''	g-like depresses	b-like depresses
''g''	g-like elevates	none

TABLE IV. Condensed chi square table comparing identification results (observed frequencies) with acoustic data (used to derive expected frequencies) (ns=not significant, *=significant at the $p=0.05$ level, and **=significant at the $p=0.01$ level).

Level of $F2$ onset										
11	**	**	**	**	**	no df	no df	no df	no df	**
10	**	**	**	**	**	no df	ns	equal	no df	**
9	**	**	**	**	**	ns	*	no df	equal?	**
8	**	**	**	**	**	ns	**	**	*	**
7	**	**	**	**	**	**	*	ns	ns	ns
6	no df	no df	no df	no df	no df	**	ns	**	**	**
5	equal	equal	no df	no df	equal	**	ns	ns	**	**
4	equal	equal	no df	no df	equal	**	**	**	**	**
3	equal	equal	equal	equal	equal	no df	**	**	ns	ns
2	equal	equal	equal	equal	equal	no df	no df	**	ns	ns
1	equal	equal	equal	equal	equal	no df	no df	**	**	ns
Vowel	ɔ	o	a	ʌ	u	æ	ɛ	ɪ	e	i

(Fig. 2), the Euclidean distance between it and each of the entire set of acoustic data points overlaid onto the identification surfaces was computed. Next, the 48 nearest-neighbor acoustic data points were identified, and then profiled in terms of their initial consonant. So, for example, for the identification stimulus at the sixth level of $F2$ onset in the vowel context series /e/, the 48 nearest neighbors from the acoustic data set included 9 bVt tokens, 25 dVt tokens, and 14 gVt tokens. This profile, 9 b's, 25 d's, and 14 g's, was then adopted as an expected set of identification frequencies for that stimulus. There was no strong theoretical basis for adopting any particular neighborhood parameters, except to preserve some notion of locality. Other choices as to the nature of the neighborhoods (size, basis, weighting by rank or distance, etc.) could have been used.

For the purposes of the chi square analysis, some modification was made to the frequencies observed in the experiment. First, ''b'' and ''w'' responses were lumped, since the expected frequencies included no ''w'' category and ''w'' and ''b'' are both labials. Second, the observed frequencies were selected from the three levels of $F3$ so that the ''b'' frequencies were taken from the experiment with a b-like $F3$, the ''d'' frequencies from the experiment with a d-like $F3$, and the ''g'' frequencies from the experiment with a g-like $F3$. The different levels of $F3$ could have been analyzed separately or simply lumped rather than selectively combined, but it was decided to make the format of the chi square analysis parallel to that of the identification surfaces presentation, in which $F3$ was for the most part deemphasized due to its relatively slight effects. Third, the ''no consonant'' responses of the identification experiment were simply ignored since there was no comparable category of expected responses and the relative frequencies of such responses in the experiment was small. Finally, the combined selected observed frequencies were rescaled to 48 to make the scale of the observed frequencies equal to that of the expected frequencies. The computations described above account for the observed frequencies in the chi square analysis not being whole numbers.

Given the expected frequencies and the observed frequencies modified as described, the chi square of each set of frequencies, for b, d, and g, for each point of locus equation space sampled in the identification experiment, was computed. A chi square using an expected frequency less than 2 was considered invalid (Guilford and Fruchter, 1978), so those cases were left out of the analysis. In consideration of this, the particular analysis protocol had to be varied by sample point. In areas of locus equation acoustic space in which only one consonant is represented, i.e., the nonoverlapped areas of [b] in back vowel space and [g] in front vowel space, the chi square had no degrees of freedom, and no interpretation was possible. However, in a substantial subset of these cases, the observed and expected frequencies are actually identical, and a statistical analysis seems superfluous. Thus, in the chi square results table (Table IV), ''no df'' means the analysis could not be applied because df=0, while ''equal'' means the analysis similarly could not be applied but also seems unnecessary due to the identity of observed and expected. For the cases in which two or three consonants were represented among the expected frequencies, i.e., there was some integration or overlap among the consonants in the acoustic neighborhood, a chi square with 1° or 2° of freedom, as appropriate, was computed and evaluated as not significant (ns), significant at the $p=0.05$ level (*), or significant at the $p=0.01$ level (**). Yate's correction for continuity was applied in those cases in which it was relevant.

As can be seen from Table II (which has an arrangement identical to Table I and parallel to Fig. 2),[2] there is generally close agreement of observed and expected frequencies in the regions of nonoverlapped acoustic data, [b] in back vowel space and [g] in front vowel space. This lends some credence to the nearest-neighbor method used to derive expected frequencies—the method yields good predictions unless there is overlap among the stops. A departure of observed from expected frequencies can be seen in virtually every region of overlap among acoustic tokens. This result is strongest in the back vowel region at higher $F2$ onset frequencies, which is the region of overlap between [d] and [g], and is also strong but relatively weaker in the front vowel region at lower $F2$ onset frequencies, which is the region of overlap between [b] and [d]. There is one anomalous set of results, for the /i/ context series at higher levels of $F2$ onset, that can be partly attributed to the breakdown of the nearest-neighbor

method. For the part of the series at the upper rightmost corner of the sample space, the same set of neighbors was sampled repeatedly simply at longer distances as the top three sample points recede from the margins of the acoustic data.

III. DISCUSSION

The results of this perception study closely resemble those obtained with only two-formant synthesis in Liberman *et al.* (1954) (cf. Fig. 1). The current study should be viewed as a cross-validation of the results of that pioneering work. Recently, convergent results were also presented by Eek and Meister (1995), although they give a very different interpetation than will be offered here. Conflicting interpretations aside, we can be reassured by the fact that three different laboratories, each using independently constructed stimuli, have separately observed the same general pattern of context-dependent consonant identifications in perception tests of burstless stimuli orthogonally varying $F2$ vowel and $F2$ onset.

The notion that a map of locus equation space somewhere in the auditory system could contribute significantly to consonant place identification is supported by the good match between the locus equation acoustic data and corresponding peaks of the identification surfaces. The darkest areas of the identification surfaces, indicating unequivocal identification of particular stops, can be thought of as analogous to partial phonological homunculi (at least as can be envisioned in these limited acoustic dimensions), while the overlaid acoustic data could represent the inputs that organized the homunculi. The correspondence between the surfaces and the acoustic data is much better for the ''b'' and ''d'' surfaces than it is for the ''g'' surface. Especially problematic is the lack of correspondence between the ''g'' surface and [g] acoustic data in back vowel contexts. This finding is interpreted here to mean that the $F2$ onset–$F2$ vowel combination is a more important cue for /b/ and /d/ than it is for /g/, at least for the current type of stimulus.

Several characteristics of the current stimuli may have contributed to the lack of ''g'' identifications in back vowel contexts. First and foremost, the stimuli were burstless, while bursts for natural [g] are generally the most prominent among the voiced stops (Smits *et al.*, 1996). Strong bursts being such a dependable correlate of [g], it is understandable that a natural-sounding burstless [g] is difficult to achieve. Even for the acoustic region yielding the most consistent ''g'' responses—front vowel contexts coupled with the higher $F2$ onset values—some subjects preferred to respond ''no consonant'' at a comparatively high rate. This general disadvantage for ''g'' is compounded in the back vowel region, where it competes with ''d'' in a region of [d]–[g] overlap in locus equation acoustic space (cf. the section to follow on the Dominance hierarchy hypothesis). Second, the stimuli were constructed with values of $F2$ onset and $F3$ onset measured at the first glottal pulse, following the standard protocol for locus equation measurements. For the g-like $F3$ condition stimuli this means that the ''velar pinch,'' the convergence of $F2$ and $F3$ onset typically found in [g] tokens, has been attenuated compared to natural tokens. Usually, the most intense overlap of the $F2$ and $F3$ onsets may be found during the [g] burst, and by the first glottal pulse the $F2$ and $F3$ transitions have already diverged somewhat, so that if samples are taken at the first glottal pulse the acute front end of the $F2-F3$ pinch will be removed. Presuming the pinch is an important cue for /g/, its attenuation in these stimuli probably contributes further to a suppression of ''g'' identifications. Finally, the values of voice onset time (VOT) and transition duration were held constant across all stimuli, and the specific values selected—a VOT of 6 ms and a transition duration of 48 ms—are briefer than those for natural [g] tokens. For example, a long VOT is frequently cited as a distinctive correlate of velars {mean [g] VOT=25 ms ($N=30$); Nossair and Zahorian (1991)}, so that a short VOT might be another factor tending to suppress ''g'' identifications.

A. Dominance hierarchy hypothesis

In interpreting consonant identification data very similar to that of this study, Eek and Meister (1995) concluded that there was no clear relationship between token-level locus equation acoustic data and the identification patterns. Such a dismissive conclusion is premature. A dominance hierarchy hypothesis is here offered to help conceptualize the relationship between the token-level acoustic data and the identification patterns for the burstless stimuli used in this study. Figure 6 presents a schematic of the hypothesis. Each outline represents a particular stop consonant's cloud of points in locus equation acoustic space. It is an abstract rendition of the combined scatterplots for [b,d,g] with the addition that the opacity of the ''clouds'' models the postulated perceptual dominance effect in regions of acoustic overlap. In combined scatterplots [d] and [b] data overlap in front vowel contexts while [d] and [g] data overlap in the back vowel region. Regarding the identification surfaces (Figs. 3–5), it can be seen that front vowel ''d'' and back vowel ''g'' are in a sense missing from the identification surfaces, similar to the results of Eek and Meister (1995) as well as the replotted data of Liberman *et al.* (1954). These identification patterns can be understood in terms of the proposed dominance hierarchy, namely, b>d and d>g. The idea is that a ''b'' identification will tend to prevail when tokens fall in the region of overlap between [b] and [d] in the front vowel region while, likewise, a ''d'' identification will tend to prevail when tokens fall in the region of overlap between [d] and [g] in the back vowel region. The stops [b] and [g] do not overlap, so their dominance relation is irrelevant.

Cross-linguistic support for the dominance hierarchy hypothesis can be obtained from the results of Eek and Meister (1995) for Estonian. A labial ''p'' was heard in 90%–100% of the cases in front vowel contexts, while ''t-syllables'' were heard as ''p-syllables'' in 70%–93% of the cases. In back vowel contexts the predominant consonant identification was alveolar ''t'' with no ''k'' responses reported (except before /a/ as also shown in our data). These Estonian data confirm the pattern of b>d in front vowel contexts and d>g in back vowel contexts for burstless stimuli.

The dominance hierarchy hypothesis framework can also be used to interpret some of the results found in Dorman *et al.* (1977), in which the vowel-dependent cue value of

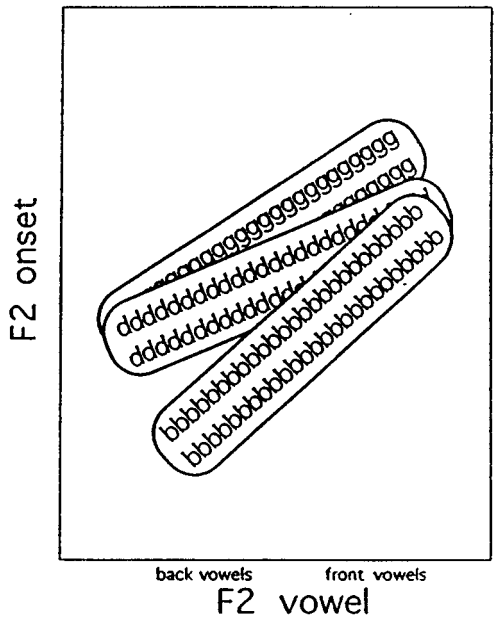


FIG. 6. Schematic of the perceptual dominance hierarchy hypothesis showing b>d in front vowel contexts and d>g in back vowel contexts and two regions of no competition—[b] in back vowel contexts and [g] in front vowel contexts.

bursts versus formant transitions for /b,d,g/ recognition was assayed. Referring, for example, to Dorman *et al.*'s Figure 4, it can be seen that the burst carries significant cue value for /d/ recognition in some front vowel contexts, but not back vowel contexts. Conversely, the burst is most valuable for /g/ recognition in back vowel contexts. Combining these observations with the patterns described by the dominance hierarchy hypothesis, one could simply conclude that the burst carries most weight in those situations in which the *F*2 vocalic transitional cue is not really distinctive—''b'' versus ''d'' in front vowel contexts and ''d'' versus ''g'' in back vowel contexts—cueing the alternative disfavored in the dominance hierarchy.

The dominance hierarchy hypothesis should be viewed as relative to stimulus properties, particularly the burstlessness of the current stimuli. The hypothesis describes the pattern of consonant identifications prevailing in nondistinctive *F*2 onset–*F*2 vowel situations when the decisive burst cue is a null. Apparently, in such cases, the stop place with the less prominent burst in natural speech prevails. Probably, if the burst is not null, it can override the default identifications described by the dominance hierarchy. Referring again to Fig. 6, a reversal of dominance relations could be envisioned as switching the position of the clouds, now the ''d'' cloud obscuring ''b'' rather than ''b'' obscuring ''d,'' for example. The dominance relation based on *F*2 onset–*F*2 vowel is b>d in front vowel contexts, but appropriate burst information should be able to override this, favoring d>b, and likewise, the dominance relation based on *F*2 onset–*F*2 vowel is d>g in back vowel contexts, but appropriate burst information should be able to override this, favoring g>d (cf. Walley and Carrell, 1983). Varying VOT or further manipulating *F*3 transition parameters might also produce dominance instabilities or reversals.

Actually, for the [b]–[d] overlap region of front vowel space, the reversal of dominance relations has already been observed. Although the subject-pooled data shows ''b'' dominating over ''d,'' at least two of the subjects clearly show the reverse relation, d>b. Despite this between-subject variability, the authors are encouraged to suppose that b>d is the prevailing dominance pattern at the group level because group data from the current experiment, Liberman *et al.* (1954), and Eek and Meister (1995) all show a dominance of ''b'' over ''d'' perception in front vowel space.

The between-subject variation noted above is one indication that the dominance of ''b'' over ''d'' in front vowel space is not as robust as the dominance of ''d'' over ''g'' perception observed in back vowel space (the latter dominance relationship showed no between-subject variation). The comparative instability of the b>d relation can also be discerned in the statistical comparisons to be discussed in the next section.

B. Chi square analysis

Significant differences between expected frequencies based on the local densities of acoustic tokens and the observed identification frequencies were found in the areas of overlap between stop consonants (Table IV). This result reinforces the conclusions previously discussed, namely that there is a dominance hierarchy in effect for the stimuli of these experiments, such that one consonant is perceived in preference to another one overlapping it in locus equation acoustic space. In contrast, regions with no overlap among stops show either exact or fairly close agreement between the expected and observed identification frequencies, although this was not always amenable to statistical test due to there sometimes being no degrees of freedom. Inspection of Table IV also indicates that the most consistent dominance effect (i.e., divergence of observed from expected frequencies) is in back vowel space in the region of [d] and [g] overlap (where d>g), while the dominance of b>d perception in front vowel space is considerably weaker, indeed at several points absent (the nonsignificant values).

One unanticipated insight afforded by the chi square analysis concerns an apparent contrast in the mode of dominance in the back vowel region versus the front vowel region. The pattern in back vowel space at higher levels of *F*2 onset is very commonly that ''d'' perceptions are more frequent than would be expected on the basis of the local density of [d] tokens, and there are as well a significant number of [d] tokens in the region, leading to a very strong dominance of ''d'' perception overall. A somewhat different pattern often occurs among the front vowels at lower levels of *F*2 onset. Here again ''d'' perceptions are often more frequent than would be expected on the basis of the local density of [d] tokens, i.e., they carry relative perceptual weight, but there are comparatively few [d] tokens in the area and an overwhelming number of [b] tokens. Thus, the somewhat weak and unstable dominance of ''b'' perception over ''d'' perception in the front vowel region might be viewed as due in part to an opposing interaction between the greater perceptual weight of [d] tokens and the far greater density of [b] tokens, while the comparatively strong, stable dominance of

''d'' perception over ''g'' perception in the back vowel region could be ascribed to the perceptual weight of [d] reinforced by its considerable density in that region.

C. Conclusion

In summary, there is ample evidence that $F2$ onset and $F2$ vowel, in combination, are significant cues for the perception of stop consonant place of articulation. These components of the speech signal are likely to be mapped together and extracted as a feature, which we have termed the vowel-normalized $F2$ transition, during speech perception. Of course, the $F2$ transition has long been considered an important cue for stop consonant place, but we are now considering a particular parametrization/coding of the $F2$ transition: in terms of its endpoints. The form of the postulated feature-extracting map could be a topographic representation of locus equation acoustic space ($F2$ vowel$\times F2$ onset).

A coherent pattern of integration of the $F2$ transition with two other sets of cues, the burst and $F3$, can be noted. In the previous discussion of the pattern of tradeoffs between $F2$ transition cues and burst cues, as described by Dorman *et al.* (1977), it was concluded that the burst carries most weight in those situations in which the $F2$ vocalic transitional cues are not really distinctive—''b'' versus ''d'' in front vowel contexts and ''d'' versus ''g'' in back vowel contexts—cueing the alternative disfavored in the postulated dominance hierarchy. The pattern of $F3$ effects in this experiment, as was summarized in Table III, is somewhat parallel to this view of the pattern of burst effects. As with the burst, $F3$ carries most weight in those situations in which the $F2$ vocalic transitional cues are not really distinctive. There is a lack of $F3$ effects in those regions in which there is a lack of overlap between the different stop places of articulation (back vowel [b] and front vowel [g]), while there are tradeoff effects between the overlapping stops in the region of their overlap ([d] and [g] in back vowel space, [b] and [d] in front vowel space). These tradeoffs are in the natural directions, with g-like $F3$ elevating ''g'' versus ''d'' identifications, and b-like $F3$ elevating ''b'' versus ''d'' identifications. Thus, it seems that one key to understanding the patterns of both burst and $F3$ effects on stop place perception is to view them in relation to the pattern of locus equation data. The way in which burst perceptual weight and $F3$ perceptual weight both correlate with the overlap versus distinctiveness of locus equation acoustic data argues strongly for a locus equation interpretation of $F2$ transitional cues.

As a closing note it is interesting that computational studies exploring properties of time-delayed neural networks also provide indirect support for $F2$ onset–$F2$ vowel playing an important role in stop place perception. When neural networks are fed sampled consonant–vowel waveforms from natural speech and instructed to classify phonetic categories by place of articulation the most effective cues for this task were reported to be $F2$ onset and $F2$ vowel frequencies (Waibel *et al.*, 1989).

ACKNOWLEDGMENTS

This research was supported by Grant No. R01 DC2014, National Institute on Deafness and Other Communication Disorders, to the second author. Portions of this manuscript are based on the Ph.D. dissertation of the first author, presented to the University of Texas at Austin. We would also like to acknowledge the constructive editorial comments of Winifred Strange, Terry Nearey, and Diana Krull.

[1]Values for the z axis (ID frequency) were derived by digitizing enlargements ($\times 5.26$) of the ID graphs from the left half of Fig. 3 in Liberman *et al.* (1954). The enlarged figures were divided into /b/ functions, /d/ functions, and /g/ functions and then separately scanned (120 dpi, 8 bit/pixel PICT files) and digitized. Estimated digitizing error was ± 1.5, i.e., less than 2.5%.

[2]A complete chi square results table is available from the second author upon request.

Bakran, J., and Mildner, V. (**1995**). ''Effect of speech rate and coarticulation strategies on the locus equation determination,'' in Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, Sweden, Vol. 1, pp. 26–29.

Bladon, R. A. W., and Lindblom, B. (**1981**). ''Modeling the judgment of vowel quality differences,'' J. Acoust. Soc. Am. **69**, 1414–1422.

Celdran, E. M., and Villalba, X. (**1995**). ''Locus equations as a metric for place of articulation in automatic speech recognition,'' in Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, Sweden, Vol. 1, pp. 30–33.

Delattre, P. C., Liberman, A. M., and Cooper, F. S. (**1955**). ''Acoustic loci and transitional cues for consonants,'' J. Acoust. Soc. Am. **27**, 769–773.

Dorman, M. F., Studdert-Kennedy, M., and Raphael, L. (**1977**). ''Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues,'' Percept. Psychophys. **22**, 109–122.

Eek, A., and Meister, E. (**1995**). ''The perception of stop consonants: Locus equations and spectral integration,'' in Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, Sweden, Vol. 1, pp. 18–21.

Fowler, C. A. (**1994**). ''Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation,'' Percept. Psychophys. **55**, 597–610.

Guilford, J. P., and Fruchter, B. (**1978**). *Fundamental Statistics in Psychology and Education* (McGraw-Hill, New York), 6th ed.

Klatt, D. H. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.

Klatt, D. H. (**1987**). ''Review of text-to-speech conversion for English,'' J. Acoust. Soc. Am. **82**, 737–793.

Krull, D. (**1988**). ''Acoustic properties as predictors of perceptual responses: A study of Swedish voiced stops,'' Phonetic Experimental Research at the Institute of Linguistics, University of Stockhom (PERILUS), Vol. VII, pp. 66–70.

Krull, D. (**1989**). ''Second formant locus patterns and consonant-vowel coarticulation in spontaneous speech.'' Phonetic Experimental Research at the Institute of Linguistics, University of Stockholm (PERILUS), Vol. X, pp. 87–108.

Liberman, A. M., and Mattingly, I. (**1985**). ''The motor theory of speech perception revised,'' Cognition **21**, 1–36.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). ''Perception of the Speech Code,'' Psychol. Rev. **74**, 431–461.

Liberman, A. M., Delattre, P. C., Cooper, F. S., and Gerstman, L. J. (**1954**). ''The role of consonant-vowel transitions in the perception of the stop and nasal consonants,'' Psychol. Mono. **68**, 1–13.

Lindblom, B. (**1963**). ''On vowel reduction,'' Report #29, The Royal Institute of Technology, Speech Transmission Laboratory, Stockholm, Sweden.

Lindblom, B., Stark, J., and Sundberg, J. (**1997**). ''From sound to vocal gesture: learning to (co)-articulate with APEX,'' to appear in FONETIK-97, Phonum **4**, Umea Universitet, Sweden.

MathWorks, Inc. (**1992**). *MATLAB Reference Guide*, Natick, MA.

Nearey, T. M., and Shammass, S. E. (**1987**). ''Formant transitions as partly distinctive invariant properties in the identification of voiced stops,'' Can. Acoust. **15**, 17–24.

Nossair, Z. B., and Zahorian, S. A. (**1991**). ''Dynamical spectral features as acoustic correlates for initial stop consonants,'' J. Acoust. Soc. Am. **89**, 2978–2991.

Smits, R., ten Bosch, L., and Collier, R. (**1996**). ''Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. I. Perception experiment,'' J. Acoust. Soc. Am. **100**, 3852–3864.

Stark, J., Lindblom, B., and Sundberg, J. (**1996**). ''APEX an articulatory synthesis model for experimental and computational studies of speech production,'' in FONETIK-96, TMH-QPSR (KTH, Stockholm), Vol. 2, pp. 45–48.

Sussman, H. M. (**1989**). ''Neural coding of relational invariance in speech: Human language analogs to the barn owl,'' Psychol. Rev. **96**, 631–642.

Sussman, H. M., Fruchter, D., and Cable, A. (**1995**). ''Locus equations derived from compensatory articulation,'' J. Acoust. Soc. Am. **97**, 3112–3124.

Sussman, H. M., Hoemeke, K., and Ahmed, F. (**1993**). ''A cross-linguistic investigation of locus equations as a relationally invariant descriptor for place of articulation,'' J. Acoust. Soc. Am. **94**, 1256–1268.

Sussman, H. M., McCaffrey, H. A., and Matthews, S. A. (**1991**). ''An investigation of locus equations as a source of relational invariance for stop place categorization,'' J. Acoust. Soc. Am. **90**, 1309–1325.

Sussman, H. M., Fruchter, D., Hilbert, J., and Sirosh, J. (**1998**). ''Linear correlates in the speech signal: The orderly output constraint,'' Target article (in press) Behavior. Brain Sci.

Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., and Lang, K. (**1989**). ''Phoneme recognition using time-delay neural networks,'' Technical Report TR-1-006, ATR Interpreting Telephony Research Laboratories, Japan, Proc. IEEE Trans. Acoust. Signal Process. **37**, 328–339.

Walley, A. C., and Carrell, T. D. (**1983**). ''Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants,'' J. Acoust. Soc. Am. **73**, 1011–1022.

The perceptual prominence of fundamental frequency peaks

C. Gussenhoven[a)]
Department of English, University of Nijmegen, Erasmusplein 1, 6525 HT Nijmegen, The Netherlands

B. H. Repp
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511-6695

A. Rietveld
Department of Language and Speech, University of Nijmegen, Erasmusplein 1, 6525 HT Nijmegen, The Netherlands

H. H. Rump
Institute for Perception Research, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

J. Terken[b)]
Institute for Perception Research, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

(Received 2 April 1996; revised 10 June 1997; accepted 1 August 1997)

Five perception experiments were conducted that investigated how the perceived prominence of $F0$ maxima in accented syllables in Dutch is affected by the variation of $F0$ minima that is supposed to relate to variation in global pitch range. The purpose of the first two experiments was to test the predictions of a model in which the reference line is directly given by an interpolation between observable $F0$ minima. The results showed that the model was inadequate, and confirmed earlier research suggesting that the reference line is calculated in a less direct way. The next three experiments investigated the role of the $F0$ of the unaccented portions of speech at the beginning (''onset'') and at the end (''offset'') of the contour, and show that only the (low) onset is used to calibrate the reference line. The results also suggest that longer onsets affect the abstract reference more than do shorter onsets. © *1997 Acoustical Society of America.* [S0001-4966(97)03711-9]

PACS numbers: 43.71.Es, 43.70.Fq [WS]

INTRODUCTION

Speakers can vary the prominence of pitch accents by varying the height of associated fundamental frequency ($F0$) maxima to express different degrees of emphasis (Liberman and Pierrehumbert, 1984; Gussenhoven and Rietveld, 1988). Likewise, listeners' prominence judgments reflect the role of $F0$ variation in relation to prominence variation (e.g., Gussenhoven and Rietveld, 1988). (We use $F0$ as a shorthand for the periodicity of the speech signal.)

Several studies have been conducted over the years that investigate the relation between $F0$ variation and perceived prominence, in order to determine the function relating $F0$ variation and perceived prominence and to develop a metric for prominence. The first relevant study was conducted by Pierrehumbert (1979). She presented listeners with utterances containing two pitch accents realized by means of $F0$ peaks, and asked them to judge which one had higher pitch. It was found that peaks later in the utterance tended to be lower than peaks earlier in the utterance when they were judged as having equal pitch; in other words, if the peaks had equal frequency, the later peak was judged to have higher pitch. This finding was explained in terms of listeners' expectations about declination, the tendency for $F0$ events later in the phrase or utterance to occur lower in the frequency range than linguistically equivalent events earlier in the domain.

In subsequent studies, beginning with Rietveld and Gussenhoven (1985), listeners were asked directly to make judgments about prominence, and several aspects of the pitch contours were varied to determine how the properties of the pitch contours feed listeners' expectations, with an emphasis on the conditions under which linguistic events later in the utterance are considered equivalent to those earlier in the utterance. These studies make it clear that listeners somehow estimate the prominence of the peak on the basis of the pitch characteristics of the contour around it, but the answer as to how listeners do this remains elusive. Among other things, it has remained unclear what the relevant characteristics of the surrounding contour are: Are they neighboring $F0$ maxima, or $F0$ minima, or both?

One opinion is that the $F0$ minima in the phrase or utterance provide a baseline which serves as a reference for the evaluation of the $F0$ maxima (Fujisaki and Hirose, 1984; 't Hart *et al.*, 1990). Due to declination, this baseline tends to be lower at the end of the phrase or utterance than at the beginning. The slope of the baseline is usually expressed in terms of semitones (st) or hertz (Hz) per s. Evidence for the relevance of baseline information in perception has been provided by Terken (1991). He employed stimuli similar to those of Pierrehumbert (1979), consisting of reiterant speech containing two accented syllables (''ma MA ma ma ma MA ma''). Artificial pitch contours were synthesized with $F0$ maxima $P1$ and $P2$ in the accented syllables, respectively.

[a)]Electronic mail: c.gussenhoven@let.kun.nl
[b)]Electronic mail: terken@ipo.tue.nl

Across stimuli, the slope of the baseline and $P1$ frequency were manipulated. Listeners went through sets of utterances, and for each set selected the utterance in which the prominence of $P2$ matched that of $P1$ most closely (''adjustment task''). Within sets, $P1$ and the slope of the baseline were held constant and only $P2$ was varied. By comparing the results of stimulus sets with and without baseline declination, it could be shown that the information contained in $F0$ minima is indeed relevant to the perceived prominence associated with $F0$ peaks. However, the results argued against the conclusion that there is a direct relation between the observable excursion size of a pitch change and its perceived prominence.

A different opinion is that $F0$ minima play no or only a minor role in determining a reference (Pierrehumbert, 1979; Liberman and Pierrehumbert, 1984; Beckman and Pierrehumbert, 1986; Ladd, 1993). These authors argue that $F0$ minima in natural utterances seldom fit nicely on a straight line, so that there would be no straightforward way for listeners to extract the baseline. Moreover, although there is evidence that the frequency of $F0$ minima varies as a function of the position of the phrase in a coherent text, (Sluijter and Terken, 1993), it appears that listeners are relatively insensitive to variations in the frequency of $F0$ minima as compared to that of $F0$ maxima (Sluijter, 1991). Hence, these authors argue that listeners determine the prominence of $F0$ peaks relative to a more abstract reference level or reference line. However, the various papers contain little concrete indication as to how listeners might determine this more abstract reference for any particular utterance.

Further problems in the modeling of the relation between $F0$ variation and prominence variation arise when the variation in overall pitch range across and within speakers is considered. The same pitch contour can be produced by speakers with different overall pitch ranges. Likewise, the same speaker can utter a sentence at different heights in his overall range. Somehow, listeners compensate for these sources of pitch variation in making prominence judgments, but it remains unclear which information listeners use to estimate the height of a pitch contour in the speaker's overall pitch range.

Summarizing, there is convincing evidence that $F0$ minima contain information that is relevant to making judgments about the prominence associated with $F0$ maxima, but it remains unclear to what extent variations in the frequency of $F0$ minima affect prominence judgments, and also whether all $F0$ minima contribute equally. The design of the experiments in Terken (1991) was limited in two ways. In the first place, a binary comparison was made between stimuli with and without baseline declination. Second, in the stimulus set in which the rate of declination was varied, it covaried with $P1$ height. Thus additional data were needed to assess the effect of variation of baseline declination and the influence of variation in peak height separately.

The remainder of the paper is structured as follows. Section I addresses in more detail the influence of $F0$ minima on the perceived prominence of $F0$ maxima. More specifically, two experiments are described which address the question to what extent variation in the rate of baseline declination affects prominence judgments, relative to the contribution of peak height variation. Among other things, the results led us to abandon the assumption that the observable baseline provides the listener with a reference line against which local $F0$ maxima are evaluated. However, they confirmed the hypothesis that both $F0$ maxima and minima are relevant to the prominence judgments. Section II describes three experiments that further explored the separate contributions of the contour beginning and the contour end to the perceived prominence of the peak(s). In Sec. III, we provide a tentative answer to the question of how the position of the contour in the overall pitch range of the speaker is estimated by the listener on the basis of the $F0$ minima around the peak, and how this estimate is used in a prominence judgment task.

I. THE RATE OF BASELINE DECLINATION

The experiments described here differ in several respects from the experiments reported in Terken (1991, 1994). First, the earlier experiments focused exclusively on male speech, and as a result, the endpoint of the declining baseline was always at a relatively low frequency. The present study attempted to rectify this situation by comparing female speech (experiment I) and male speech (experiment II). Second, a simple meaningful sentence was used instead of reiterant speech. This was expected to help subjects maintain a linguistic frame of reference, at the risk of introducing other confounding variables. Third, in experiment I two tasks were compared: a single-trial relative prominence judgment task (henceforth, ''judgment task'') and a ''quasi-adjustment task'', as used by Terken (1991). The judgment task, in particular, was thought to discourage a literal pitch matching strategy. Finally, a group of relatively naive subjects (i.e., not including any speech researchers) was used. In Terken's (1991) study there seemed to be a tendency for the most experienced listeners to show the smallest effects. The segmental homogeneity of his stimuli, his adjustment paradigm, and his listeners' experience may all have encouraged an analytic listening strategy, counteracting the adoption of the ''linguistic mode'' most favorable to prominence judgments.

A. Experiment I

1. Method

A female Dutch speaker spoke the sentence ''Amanda gaat naar Malta'' (''Amanda goes to Malta''), with pitch accents on –man– and –mal–, in a neutral fashion without special emphasis on ''Amanda'' or ''Malta.'' The speech was recorded in a recording studio using high-quality equipment. The utterance was input to a computer, digitized at 20 kHz, and low-pass filtered at 10 kHz. It was then LPC coded (24 coefficients, 10 ms frames), and $F0$ was determined using the method of subharmonic summation (Hermes, 1988). A stylized $F0$ contour was fitted to the utterance following the procedures described in 't Hart *et al.* (1990), to obtain an estimate of the baseline declination. This gave a declination rate of 2.4 st/s for the baseline, and an end frequency of 164 Hz. Phonetic segment boundaries were measured using a waveform editing program. The total utterance was 1635 ms in duration; the critical /man/ and /mal/ syllables had dura-

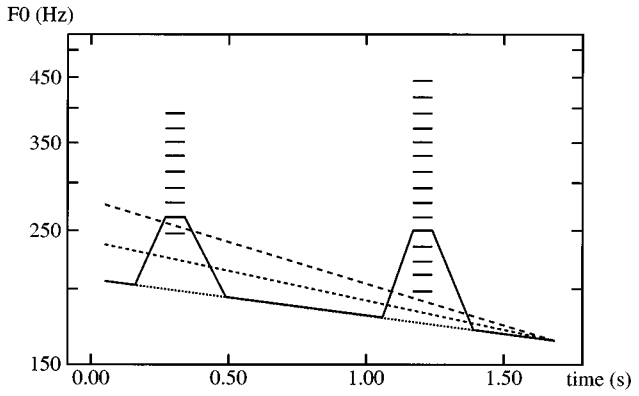


FIG. 1. Schematic illustration of stimulus parameters in experiment I (female speaker). The solid line represents an actual $F0$ contour for the sentence *Amanda gaat naar Malta* (''Amanda goes to Malta'') with pitch accents on $-MAN-$ and $-MAL-$. The dashed lines represent the declination lines for different declination conditions. Also the ranges for the peaks of the pitch accents on $-MAN-$ and $-MAL-$ are indicated.

tions of 291 ms (/m/: 87 ms; /an/: 204 ms) and 268 ms (/m/: 106 ms; /al/: 162 ms), respectively. There were three voiceless gaps: for /χ/ and /t/ in ''gaat,'' and for /t/ in ''ta''; the last one occurred during the fall of $P2$.

The basic $F0$ contour used in the experiment consisted of a baseline with two rise-and-fall combinations superimposed on it, for $P1$ and $P2$, respectively (cf. Fig. 1). A set of 105 $F0$ contours was constructed by combining 3 declination rates, 5 $P1$ heights, and 7 $P2$ heights. The three declination rates were 2.4 (original), 3.9, and 5.5 st/s, with a fixed end frequency of 164 Hz. The rises began 30 ms before the vowel onsets of the accented syllables and lasted 110 ms; the falls began 50 ms after vowel offset and lasted 150 ms; in between the rise and fall there was a flat top lasting 70 ms (these values were inspired by the properties of the original contour). Different $P1$ and $P2$ values were produced by manipulating the size of the rise-and-fall combinations (i.e., the frequency interval between onset and offset of the rise and the fall), while keeping the duration constant.

The combinations of declination rate, $P1$ height, and $P2$ height were not strictly orthogonal because the range of $P1$ heights necessarily had to covary to some extent with the height of the baseline, and the range of $P2$ heights had to be approximately centered around each $P1$ height to yield an unbiased estimate of the point of equal prominence. The precise stimulus design is shown in Table I. As can be seen in the table, 9 $P1$ values were assigned to the 3 baselines in 3 overlapping sets of 5 (giving 15 combinations of $P1$ and declination rate), and 15 $P2$ values were assigned to the 9 $P1$ values in 9 overlapping sets of 7 (giving 105 possible combinations of $P1$, declination rate and $P2$). $P1$ values ranged from 247 to 392 Hz in steps of approximately 1 st, and $P2$ values ranged from 198 to 444 Hz, also with a 1-st step size. These parameters are illustrated in Fig. 1.

For the judgment task, the 105 stimuli and 5 fillers were recorded on digital tape five times, each time in a different random order. The fillers represented successive steps in the process of going from the natural contour to the basic artificial contour. Due to space constraints the results for the fillers will not be presented here. Each block of 110 stimuli lasted about 10 min. The interstimulus interval was about 4 s. A longer interval occurred after the 35th and the 70th trials, and a still longer interval in between blocks. The five test blocks were preceded by a practice sequence containing, in random order, the 30 stimuli in which $P2$ had one of the extreme values in its 7 value range (i.e., all the first and last stimuli in the rows of Table I). Subjects were tested in two groups in a quiet listening room. The test tape was presented over earphones at a comfortable loudness level. Subjects were asked to judge for each stimulus which of the two words, ''Amanda'' or ''Malta,'' was given more emphasis by the speaker. (The instructions pointed out that this was the same as judging the relative prominence of the accented syllables /man/ and /mal/.) The responses were made by writing

TABLE I. Stimulus design, showing the range of $P2$ values for each $P1$ value at different declination rates, and percentages of ''1'' responses (where ''1'' means that $P1$ was judged to be more prominent) to each stimulus in the judgment task (Experiment I).

Decl.	$P1$	$P2$ (Hz)														
(st/s)	(Hz)	198	211	222	235	250	263	278	294	313	333	351	370	392	417	444
2.4	247	94	77	80	66	54	26	11								
	263		97	91	71	69	31	20	9							
	278			86	86	71	63	29	9	3						
	294				97	77	46	43	34	0	9					
	313					91	74	54	43	17	3	6				
3.7	278			94	97	80	63	49	17	11						
	294				89	80	77	51	34	6	9					
	313					94	91	77	46	26	3	0				
	333						91	83	66	31	11	3	14			
	351							97	77	49	34	6	3	3		
4.9	313					91	83	89	60	14	26	6				
	333						94	91	63	66	20	9	3			
	351							91	80	57	37	20	9	6		
	370								94	80	63	26	6	6	0	
	392									89	80	54	17	3	0	0

''1'' or ''2'' on a prepared answer sheet. A forced choice was required on every trial. The experiment took a little over one hour, including instructions and a short break after the third block. The subjects were seven members of the research staff at IPO, four women and three men, all native speakers of Dutch, who volunteered to participate.[1] None of them was involved in speech research.

About two months later, the same subjects took part in the quasi-adjustment task, which lasted about 30 min. Subjects were tested individually in a quiet room with instructions similar to those in the judgment task. They sat in front of a computer screen and listened to the stimuli over earphones. Subjects worked through a series of runs. Within a given run, $P1$ and declination were fixed and only $P2$ varied; that is, a given run used the seven stimuli that had the same combination of $P1$ and declination rate and differed only with respect to $P2$ (one of the rows in Table I). In each run they ''adjusted'' the frequency of $P2$, by selecting the stimulus in the run in which the prominence of $P2$ matched the prominence of $P1$ most closely. The subject was first presented with the stimulus having the lowest (or highest) $P2$, and could then press any one of the digit keys between 1 and 9 on the computer keyboard to hear the next stimulus. Pressing key 5 corresponded to ''replay the last stimulus,'' pressing keys below 5 corresponded to selecting stimuli with lower $P2$ if available (the farther away from 5 the larger the difference from the current stimulus), and pressing keys above 5 corresponded to selecting stimuli with higher $P2$ than the current stimulus (again, the farther away from 5 the larger the difference). They completed 60 adjustment runs, 4 for each of the 15 combinations of $P1$ height and declination rate, arranged in 4 blocks of 15 runs each. The adjustments were subsequently averaged over the 4 replications.

2. Results and Discussion

For the 105 test stimuli in the judgment task the number of ''1'' responses was tallied, added up across subjects, and converted into percentages. These percentages are shown in Table I. Although some subjects were not totally confident even at the extremes of the $P2$ ranges, an orderly progression from ''1'' to ''2'' judgments can be seen in each of the 15 rows of the table.

Probit analysis (Finney, 1971) was used to estimate the 50% crossover point from ''1'' to ''2'' responses within each of the 15 series of 7 $P2$ values, individually for each subject. In one anomalous instance, the first two percentages of a nonmonotonic response function were ignored in order to get a reasonable estimate. These estimates represent the points of equal judged prominence of $P1$ and $P2$. Figure 2(a) shows these estimates, averaged across the seven subjects. Figure 3(a) gives the corresponding mean values for the adjustment task, which yielded such estimates directly. The results of the two experiments are quite similar. Despite some variability at the individual level, the overall results are remarkably orderly.

The average data show the expected $P1-P2$ difference: in all cases $P2$ was lower than $P1$ when it was judged to be equal in prominence.[2] All seven subjects showed this overall effect in both tasks.

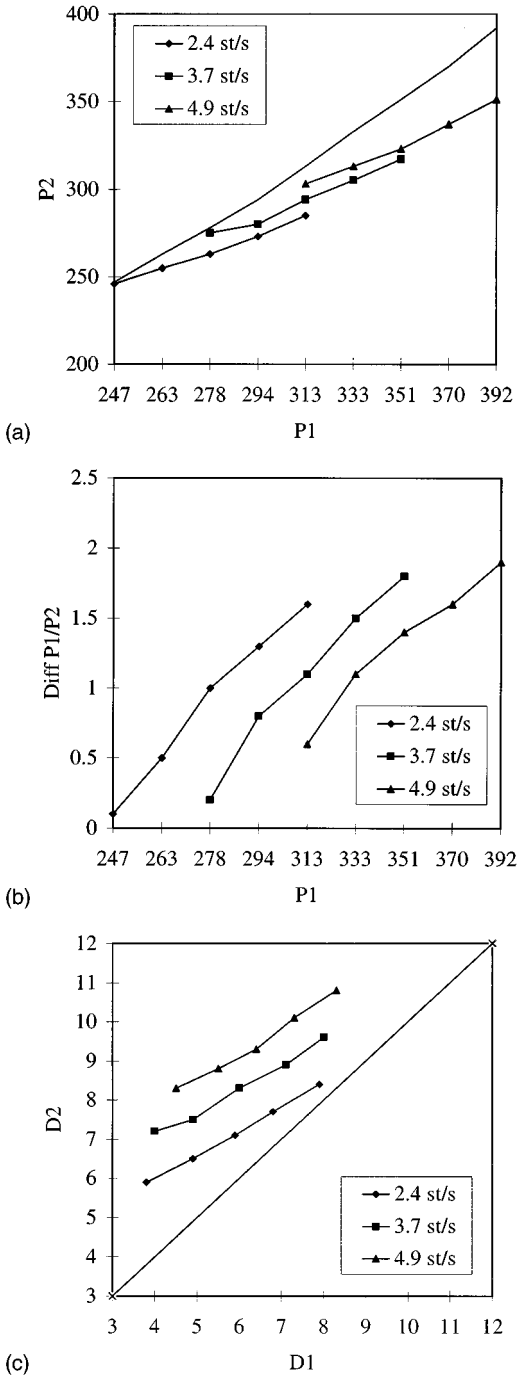


FIG. 2. Mean values of $P2$ giving the same prominence as $P1$, as a function of $P1$ height, for different declination rates (experiment I, judgment task). (a) Peak values for $P2$ against $P1$; the solid diagonal-like line represents the $P2=P1$ curve; (b) Differences between peak values for $P1$ and $P2$ (Diff $P1/P2$, in semitones) against $P1$ peak values; (c) Excursion size $D2$ (distance between peak $P2$ and the interpolated baseline) against excursion size $D1$.

Figures 2 and 3 also show that the discrepancy between $P1$ and $P2$ (Diff $P1/P2$) increased with $P1$ height [this was true both when expressed in Hz and st, see Figs. 2(b) and 3(b)]. The consistency of this effect was assessed by fitting straight lines to the five equal-prominence $P2$ values obtained in each declination condition, separately for each subject (there were no systematic nonlinearities in any subject's data), and by examining the slopes of these lines. Nearly all

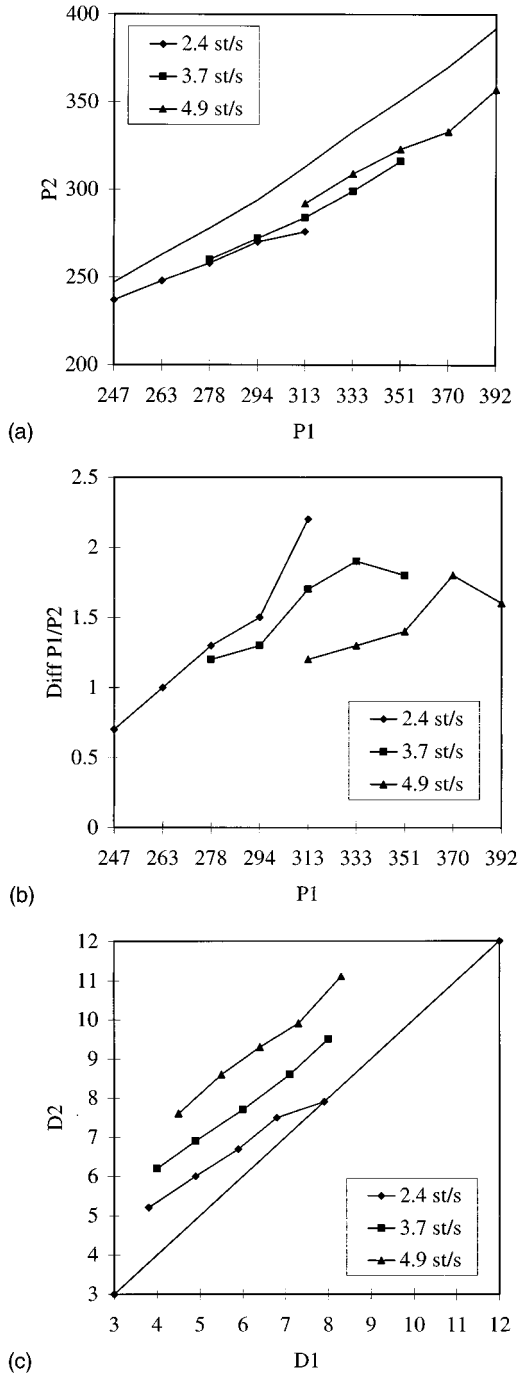


FIG. 3. Mean values of $P2$ giving the same prominence as $P1$, as a function of $P1$ height, for different declination rates (experiment I, adjustment task). (a) Peak values for $P2$ against $P1$; the solid diagonal-like line represents the $P2=P1$ curve; (b) Differences between peak values for $P1$ and $P2$ (Diff $P1/P2$, in semitones) against $P1$ peak values; (c) Excursion size $D2$ (distance between peak $P2$ and the interpolated baseline) against excursion size $D1$.

were less than 1. The increase in the $P1-P2$ difference with $P1$ height was statistically reliable, as shown by a sign test ($z=5.71$, $p<0.001$). The slopes tended to be steeper in the adjustment task, but this task difference fell short of significance [$F(1,6)=5.6$, $p<0.06$].

Third, the ANOVA on the slopes of the lines fitted to the data points revealed a significant main effect of declination rate [$F(2,12)=4.3$, $p<0.04$]: There was a significant trend for the slopes to get steeper as the declination rate increased. In other words, the increase in the $P1-P2$ difference with $P1$ height was more pronounced with low than with high declination rate. This effect did not interact with task either. However, the data suggest that the slope differences were essentially restricted to the adjustment task.

Finally, Figs. 2(c) and 3(c) show that the excursion sizes $D1$ and $D2$, i.e., the distances between the peaks and the interpolated baselines, varied at different rates for $P1$ and $P2$. This confirms earlier findings that equal prominence is not simply a matter of equal excursion sizes.

Our primary interest is in the effect of rate of declination on the relation between $P1$ and $P2$. This effect was relatively small, as may be seen in Figs. 2(b) and 3(b) by connecting equal $P1$'s: Diff $P1/P2$ decreased slightly when the baseline became steeper. That is, for a given $P1$, $P2$ should be slightly higher in stimuli with steeper declination than in stimuli with less steep declination, in order to maintain equal prominence. The significance of this effect was tested in three ANOVAs on the $P2$ estimates: two for the two sets of $P1$ values that were shared by two declination conditions (278, 294, 313 Hz, and 313, 333, 351 Hz, respectively), and one for the single $P1$ value that was shared by all three (313 Hz). The declination effect was significant in all three analyses [low versus medium declination rate: $F(1,6)=13.4$, $p<0.02$; medium versus high declination rate: $F(1,6)=56.9$, $p<0.0004$; all three: $F(2,12)=17.6$, $p<0.0004$], and there was no significant interaction with task or with $P1$ height. Of course, there were highly significant main effects of $P1$ height in the first two analyses, but the interactions of that effect with task were nonsignificant. The results will be further discussed in combination with those of experiment II.

B. Experiment II

1. Methods

A male Dutch speaker spoke the same sentence as in experiment I, ''Amanda gaat naar Malta.'' The speech was digitized at 20 kHz and low-pass filtered at 10 kHz. The utterance was LPC coded (32 coefficients, 10-ms frames), and the $F0$ contour was determined. This contour was then subjected to stylization, which was straightforward as the unaccented syllables could be fit easily by a single straight (i.e., exponentially) declining line. The declination rate was 3.9 st/s, which is almost exactly the rate predicted by the Dutch synthesis model (3.8 st/s), and the end frequency was 102 Hz. In contrast to the source utterance of experiment I, $P2$ was much lower than $P1$ in the natural utterance used to make the stimuli for experiment II. However, this was considered irrelevant, as $F0$ was changed in the experimental stimuli. The total utterance was about 1410 ms in duration; the critical /man/ and /mal/ syllables had durations of about 230 ms (/m/: 80 ms; /an/: 150 ms) and 270 ms (/m/: 80 ms; /al/: 190 ms), respectively.

As in experiment I, a set of 150 $F0$ contours was constructed by combining 3 declination rates, 5 $P1$ heights, and 7 $P2$ heights, selected from a range of 9 $P1$ and 15 $P2$ values. The two accent peaks were modeled on those of the

TABLE II. Range of $P2$ values for each $P1$ value in experiment II.

Decl. (st/s)	$P1$ (Hz)	$P2$ (Hz)														
2.4	150	119	126	133	141	150	159	168								
	159		126	133	141	150	159	168	177							
	168			133	141	150	159	168	177	189						
	177				141	150	159	168	177	189	200					
	189					150	159	168	177	189	200	211				
3.9	168			133	141	150	159	168	177	189						
	177				141	150	159	168	177	189	200					
	189					150	159	168	177	189	200	211				
	200						159	168	177	189	200	211	225			
	211							168	177	189	200	211	225	238		
5.5	189					150	159	168	177	189	200	211				
	200						159	168	177	189	200	211	225			
	211							168	177	189	200	211	225	238		
	225								177	189	200	211	225	238	250	
	238									189	200	211	225	238	250	267

female utterance of experiment I, and were similarly aligned with the segmental structure: the durations of rise, top, and fall portions were 110, 70, and 150 ms, respectively, for both peaks, with the rise starting 30 ms prior to vowel onset. The $F0$ of the top portion declined at the baseline rate, so that the maximum $F0$ occurred at the end of the rise. The temporal distance between $P1$ and $P2$ was 650 ms. $P1$ heights ranged from 150 to 238 Hz, and $P2$ heights from 119 to 267 Hz, in steps of approximately 1 st. The precise $P1$ and $P2$ combinations are presented in Table II. The three declination rates chosen were 2.4, 3.9 (original), and 5.5 st/s. The basic set up of the stimulus set is similar to that of experiment I (cf. Fig. 1).

Only an adjustment task was used in experiment II, as the two tasks in experiment I had yielded rather similar results. Instructions and testing procedure were identical to those in the adjustment task of experiment I. The subjects were ten members of the research staff at IPO, five women and five men, all native speakers of Dutch, who volunteered to participate. None of them was involved in speech research, and none had participated in experiment I.

2. Results

The results are shown in Fig. 4. As in experiment Ib, a strong effect of initial $P2$ height was noted: When the first stimulus in an experimental run contained the highest value of $P2$, adjustments were much higher than when the run started with the lowest $P2$ value. The data presented represent an average over both types of adjustment runs.

Again, $P2$ was always adjusted to a lower value than $P1$, which replicates the basic $P2<P1$ effect [cf. Fig. 4(a)]. The effect was shown throughout by all but two subjects, who adjusted $P2$ to values higher than $P1$ when $P1$ was above 200 Hz. Furthermore, the majority of the subjects increased the $P1-P2$ difference (expressed in st) with increasing $P1$, as can be seen by connecting equal $P1$'s in Fig. 4(b), although the effect was smaller and less orderly than in experiment I. A combined analysis with the slopes of the regression lines of individual subjects for the adjustment data of experiment I did not yield a significant difference between experiments [$F(1,15)=2.20$, $p=0.16$], and overall the slopes were again less than 1 [$F(1,15)=17.01$, $p<0.001$]. Unlike that found in experiment I, the difference in slopes across declination conditions was not significant [$F(2,18)=1.30$, $p=0.30$].

As in experiment I, the effect of principal interest, that of rate of declination on the relation between $P1$ and $P2$, was relatively small [cf. Fig. 4(a)], and it was only marginally significant in ANOVAs on the $P2$ responses to the $P1$ values shared by the different declination conditions [low versus medium: $F(1,9)=4.86$, $p<0.06$; medium versus high: $F(1,9)=7.77$, $p<0.03$; all three: $F(2,18)=3.10$, $p<0.07$]. Also, as in experiment I, for each separate declination condition, the distance between $P1$ and the interpolated baseline ($D1$) covered a different range of values than the distance between $P2$ and the interpolated baseline ($D2$) [cf. Fig. 4(c)].

3. Discussion

First, the results replicate the well-known finding that $P2$ should be lower than $P1$ to give an impression of equal prominence, i.e., the declination effect. Second, the results confirm earlier findings that the $P1-P2$ difference required for equal prominence increases as $P1$ increases, although the effect is smaller here than in Terken (1991), which matched the set-up of the current experiments most closely. Third, the effect of declination, which is of main interest here, is relatively small: For a given $P1$, a higher declination rate requires a small increase in $P2$ frequency to maintain equal prominence.

The finding that an increase in the rate of declination requires a small increase in $P2$ frequency to maintain equal prominence is in contrast with the findings of Terken (1991), where the effect of the presence of declination was to *lower* $P2$ relative to the ''no declination''-condition. One possible explanation for this discrepancy is in terms of the different methods used. The current experiments differed from those in Terken (1991) with respect to both stimulus properties and

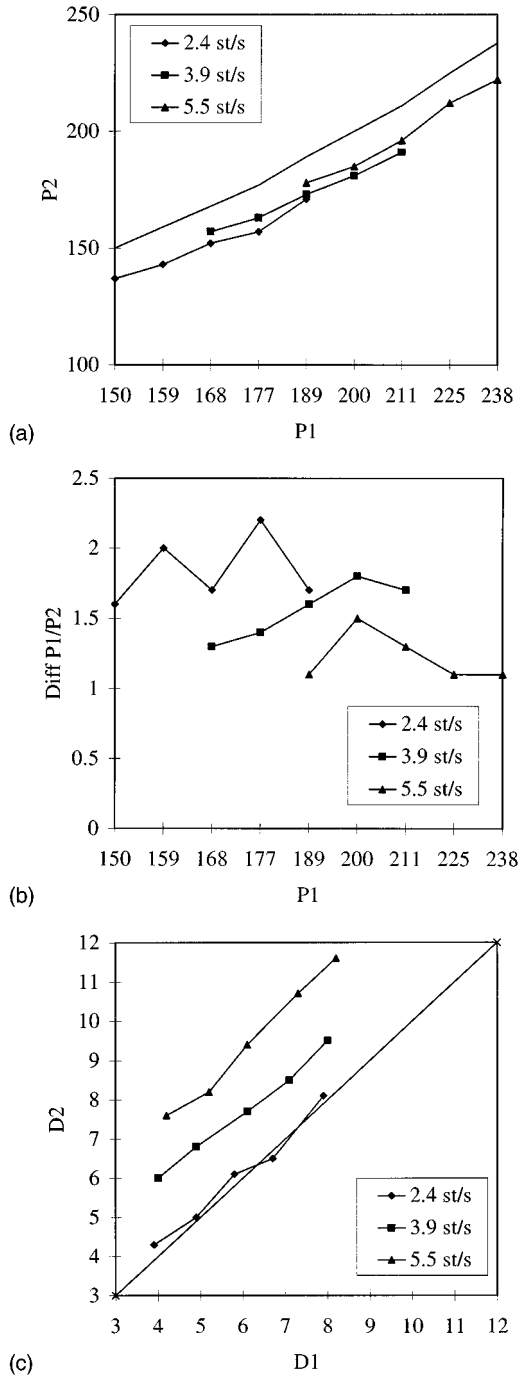


FIG. 4. Mean values of $P2$ giving the same prominence as $P1$, as a function of $P1$ height, for different declination rates (experiment II, adjustment task). (a) Peak values for $P2$ against $P1$; the solid diagonal-like line represents the $P2=P1$ curve; (b) Differences between peak values for $P1$ and $P2$ (Diff $P1/P2$, in semitones) against $P1$ peak values; (c) Excursion size $D2$ (distance between peak $P2$ and the interpolated baseline) against excursion size $D1$.

presentation methods. In particular, whereas in the current experiments the stimuli representing the different declination conditions were presented in a random order in the same experiment, in the experiments described in Terken (1991) they were presented in separate experiments. This may have induced a different, more analytic mode of listening in the ''flat baseline'' condition than was used for the more natural ''declination'' condition, and $P2$ may have been adjusted to higher values in the ''flat baseline'' condition than would have been the case under a more synthetic listening mode. Thus the two conditions in Terken (1991) may have provided qualitatively different outcomes, so that the comparison between the conditions of Terken (1991) would not be not relevant to our present concerns.

The results of experiment I and experiment II can be summarized as follows. If $P1$ is raised while keeping the baseline constant (so that the distance between the peak and the baseline increases), the difference between $P1$ and $P2$ that is required for equal prominence increases (a finding that was obtained only in experiment I). Second, if the declination becomes steeper while keeping $P1$ fixed (so that the distance between the peak and the baseline decreases), the $P1-P2$ difference required for equal prominence decreases. Thus, there would appear to be a tendency for the difference between $P1$ and $P2$ to be proportionally related to the distance between $P1$ and the baseline, regardless of the position of the contour in the overall range of the speaker.

Our results argue against the hypothesis that prominence is related in a direct way to the distance between the peaks and the overt baseline, a conclusion that concurs with those of Pierrehumbert (1980:128) and Ladd (1993), who assume the reference is abstract. At the same time, the results provide evidence for the relevance of the $F0$ minima in the contour to the perceived prominence of $F0$ peaks. Therefore, we need to explore how listeners construct an abstract reference line on the basis of the physical characteristics of the contour. To this end, the role of $F0$ minima was investigated in a more piecemeal fashion. These investigations are described in Sec. II.

II. ANCHORING THE REFERENCE LINE

A logical step to take, one that emerges from the conclusions of Sec. I, is to test for the effect of the $F0$ minima at different locations in the contour and try to establish their effects independently of each other. In experiment III, we tested for the effect of the final low pitch. In experiment IV, we did the same for the beginning of the contour. Finally, in experiment V we again looked at the influence of the beginning of the contour in order to replicate the results of experiment IV. In addition, with experiment V we returned to the issue of the declination effect, and investigated it as a function of the length of the utterance before the pitch peak.

A. Experiment III: Offset height

The purpose of experiment III was to test the relevance of the low end point of the contour to the perceived prominence of the peak or peaks in it. While the results of experiments I and II suggest that a raising of the beginning of the contour reduces the prominence of peaks, no information has been obtained about a possible effect of the contour end. The depth of the downward trajectory at the end of a falling contour might well determine the perceived prominence of the final peak, or of all peaks, in the contour. Also, this effect may depend on how close the final peak is to the end of the contour.

1. Methods

A male speaker recorded the sentence *Dat geblaat de hele dag van die schapen daar* (''that bleating all day of those sheep there'') in a neutral, somewhat monotonous fashion, without placing particular emphasis on any of the words. The whole utterance had a duration of 2215 ms; the word *geblaat* had a duration of 480 ms, and the word *schapen* 440 ms. The syllable *blaat* had a duration of 294 ms ([a:]: 132 ms); the syllable *scha* had a duration of 293 ms ([a:]: 142 ms). The utterance was digitized at a sampling rate of 10 kHz, with the LP filter set at 5 kHz, LPC analyzed (18 coefficients, 10-ms frames). Our utterance served as the basis of three syntactically exclamatory sentences. After splicing off the portion of the speech file corresponding to *van die schapen daar*, the shortened speech file was used to create a set of one-peak stimuli, while the original speech file was used to create a set of one-peak stimuli as well as a set of two-peak stimuli. These three carrier sentences are listed below, with their durations; the capitalized syllables indicate the locations of the $F0$ peaks in the artificial contours.

(1) Dat geBLAAT de hele dag! (1255 ms) ''that bleating all day''
(2) Dat geBLAAT de hele dag van die schapen daar! (2215 ms)
(3) Dat geBLAAT de hele dag van die SCHAPen daar! (2215 ms) ''that bleating the whole day of those sheep there''

The artificial pitch contours contained a level baseline of 100 Hz. Peaks had flanks of 100 ms, and were aligned such that the $F0$ maxima coincided with the midpoint in the accented vowel. Three values for the peaks were used: 140, 150, and 160 Hz. In the two-peak contour the values of $P1$ and $P2$ were the same. After the last peak, the pitch fell to a low point, to which we refer as the ''offset.'' The offset was also varied in three steps, 95, 87, and 80 Hz. Offsets and peaks were crossed, which led to nine stimuli per sentence (see Fig. 5).

Two audiotapes were prepared on each of which the 27 stimuli appeared three times in a random order. The tapes were presented through headphones in a language laboratory to 20 subjects, about half of whom listened to one version and the rest to the second version. All our subjects were students at the University of Nijmegen, who were paid a small fee for their services. They were given a written instruction sheet along with the response form.

Their task was to rate one of the accented syllables in each stimulus for general prominence. On their score sheets, the corresponding sentence was printed with the syllable to be scored printed in capitals. This syllable was always associated with a pitch peak in the stimuli. For each stimulus, they gave their judgment by putting a tick on a 100-mm scale, which had the words ''little emphasis'' printed on the left and the words ''much emphasis'' on the right. After 15 stimuli an anchor stimulus was included with one accented syllable. The prominence of this anchor stimulus was shown on the score sheets as being on the mid-point of the scale. The stimuli were preceded by ten test trials; the first two and the last two stimuli were fillers. The stimuli were presented in blocks of ten, with an ISI of 6 s, and intervals of 10 s between blocks.

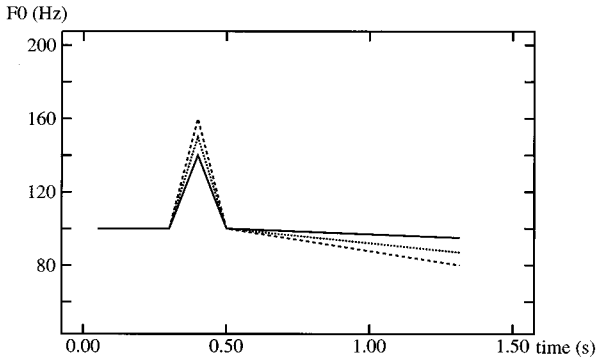


FIG. 5. Schematic representation of $F0$ contours for stimuli in experiment III. The solid line represents an actual contour for the sentence *dat geblaat de hele dag* (''that bleating all day'') with a pitch accent on *–BLAAT–*. Dashed lines represent alternative $F0$ peaks for the pitch accent and $F0$ trajectories for the offset.

2. Results

The three sentences differed in the length of the post-peak stretch, i.e., the part following the target accent on ''geblaat,'' and in the presence in the post-peak stretch of a pitch peak on ''schapen.'' Figure 6 gives the mean prominence ratings for the different combinations of offset and peak height for the three sentences. As can be seen, for all sentences the perceived prominence increased as the peaks increased in fundamental frequency. However, for no sentence does there appear to be an effect of offset height. This is confirmed by two analyses of variance (repeated measures). One analysis was performed on the scores for *geblaat* in sentences 1 and 2, averaged over the three replications, with the fixed within-subjects factors ''Peak Height,'' ''Offset Height,'' and ''Distance to Offset.'' The other analysis was performed on the scores of sentences 2 and 3, averaged over the three replications, with the fixed within-subjects factors ''Peak Height,'' ''Offset Height,'' and ''Number of Peaks.'' In both analyses, only the factor ''Peak Height'' was significant: $F(2,60)=38.67$ and $F(2,60)=27.87$, respectively; Huynh–Feldt corrected p value <0.05.

3. Discussion

The perceived prominence of an $F0$ peak appears to be independent of the pitch of the final low end point of the contour. This finding may seem surprising, because the final low point would appear to be such an obvious candidate as a cue to the speaker's overall pitch range. It has been found to be a relatively invariant value for speakers in laboratory speech (Liberman and Pierrehumbert, 1984), and so might well have served as an anchor point for the speaker's pitch range (assuming it is calculated afresh for each stimulus). The fact that it is not rules out models of prominence perception which depend on a reference line—whether overt or abstract, whether horizontal or descending—which is determined by the contour end. With hindsight, however, it is not implausible that the final low pitch should not be used as an

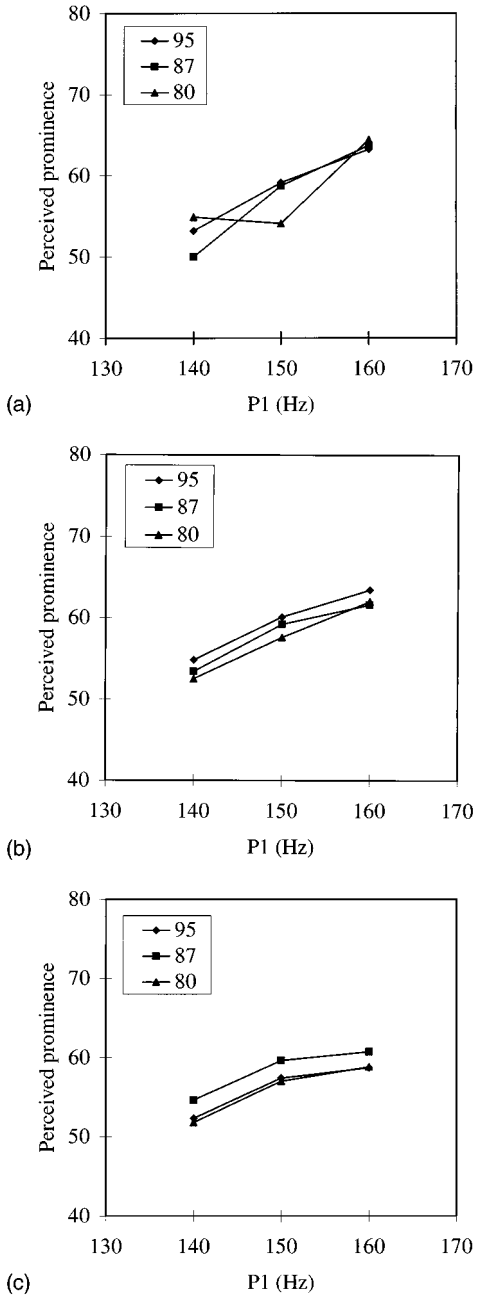


FIG. 6. Mean perceived prominence ratings for *geBLAAT* (''bleating'') on a 100-mm scale as a function of Peak Height and Offset frequency (experiment III). (a) Sentence 1, with short post-peak stretch: *dat geBLAAT de hele dag*; (b) Sentence 2 with long-post-peak stretch: *dat geBLAAT de hele dag van die schapen daar*; (c) Sentence 3 with long post-peak stretch containing an accent on *SCHApen* (''sheep''): *dat geBLAAT de hele dag van die SCHApen daar*.

anchor point for pitch range. Final low pitch serves as a cue to finality. Although the distinction between final phrases and nonfinal phrases largely correlates with contour type, with a boundary high tone signalling nonfinality and a boundary low tone signalling finality (Geluykens and Swerts, 1994), at least some of the variation found by Swerts *et al.* (1994) between final and nonfinal phrases in spontaneous dialogue could be attributed to the occurrence of a ''half-fall'' versus a fall to the bottom of the pitch range. Moreover, there is evidence that the final fall is used by speakers to signal the end of their turn at talk, with deeper falls being more likely to lead to turn taking by the listener (Beattie *et al.*, 1982). Thus the results of experiment III can be explained by assuming that the final low pitch is itself evaluated on the basis of the listener's knowledge of the speaker's pitch range. Obviously, in this scenario, the overall pitch range needs to be known to the listener before the degree of finality of the speaker's utterance can be evaluated on the basis of the final low pitch.

An additional reason for assuming that the final low end point is a poor anchor for the reference line is that it varies much less under changes in register, i.e., the position of the contour in the speaker's overall range, than other aspects of the contour (Ladd and Terken, 1995; Mozziconacci, 1995). Thus information situated elsewhere in the contour would seem preferable to anchor the reference line.

B. Experiment IV: Onset height

The purpose of experiment IV was to test the relevance of the initial pitch of the contour to the perceived prominence of the peak or peaks in it. We use the term ''onset'' to refer to the stretch between the contour beginning and the rise towards the accent peak. The pitch of the onset might serve as a cue to the contour's reference line, with higher onsets signalling higher reference lines, which might consequently lead to less perceived prominence of the peak. Since any effect of the onset might depend on its length, we also varied the distance between the contour beginning and the peak.

1. Methods

We used the following sentences, obtained from the original utterance used in experiment III, as the basis of our stimuli:

(1) Dat geBLAAT de hele dag! (1255 ms) ''that BLEATING all day''
(2) Dat geblaat de hele dag van die SCHAPen daar! (2215 ms) ''that bleating all day of those SHEEP there''

The peaks were varied in three steps, with peak values of 140, 150, and 160 Hz, as before. These steps were crossed with four onset heights, 100, 108, 116, and 122 Hz. After the last peak, all contours descended to a low pitch of 87 Hz. The set-up is shown schematically in Fig. 7. The same 20 raters who participated in experiment III, were asked to judge the prominence of *geblaat* in the short sentence and *schapen* in the long sentence. In other respects, methods were similar to those in experiment III.

2. Results

Figure 8 gives the average ratings for the different combinations of onset and peak height for the different sentences. An analysis of variance (repeated measures) was performed on the data, with the fixed within-subject factors ''Peak Height,'' ''Onset Height,'' and ''Sentence.'' ''Peak Height'' and ''Sentence'' were significant: $F(2,38)=50.42$, $p<0.01$ and $F(1,19)=35.04$, $p<0.01$ (Huynh–Feldt corrected). Thus as expected, variation in peak height affected perceived prominence. In addition, the peak on *ge*BLAAT in

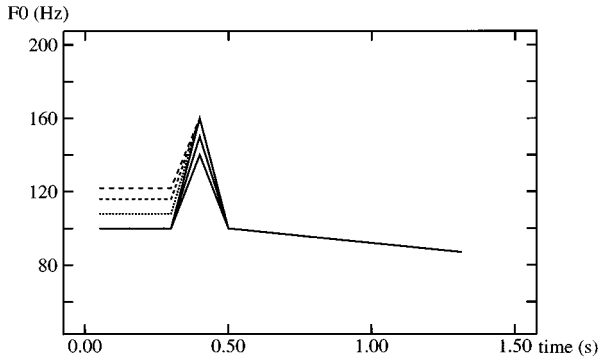


FIG. 7. Schematic representation of $F0$ contours for stimuli in experiments IV and V. The solid line represents the contour for the sentence *dat geblaat de hele dag* (''that bleating all day'') with the three peaks for the pitch accent on *–BLAAT–*. Dashed lines represent the four trajectories for the onset with which the three peaks were crossed.

the sentence with the short onset was on average less prominent than that on SCHA*pen* in the sentence with the long onset [60.3 vs 71.8, cf. Fig. 8(a) and (b), respectively]. This main effect may point to an effect of onset length, but it may also be due to influences not controlled for in the experiment, like the relative durations or loudness levels of the syllables *blaat* and *scha–*. What is of interest in these data is the interaction between ''Sentence'' and ''Onset Height:''

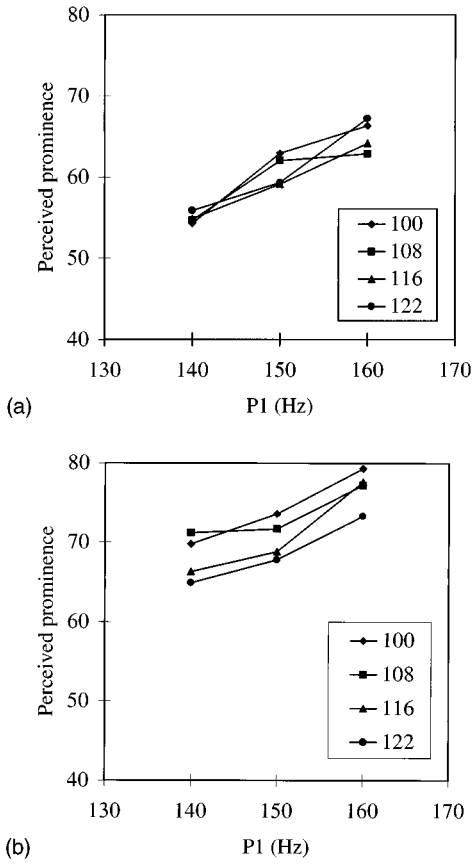


FIG. 8. Mean perceived prominence ratings on a 100-mm scale for *geBLAAT* (''bleating'') as a function of Peak Height and Onset Height, for different sentence lengths. (a) Short sentence, with short onset (Sentence 1): *dat geBLAAT de hele dag*; (b) Long sentence, with long onset (Sentence 2): *dat geblaat de hele dag van die SCHApen daar*.

$F(3,57)=3.95$, $p=0.018$. The perceived prominence of the peak after the long onset decreased when the onset was raised. That is, the higher the onset, the less prominence the raters perceived on the peak. In the utterance with the short onset, there is no such effect of onset height.

3. Discussion

The results of Experiment IV show that the height of the onset may affect the perceived prominence of the peak. The way in which the perceived prominence depends on the onset, with higher onsets reducing the prominence of the peak, suggests that the listener uses the onset as a cue to the contour's reference line. However, it would not appear as if it is literally the starting point of the contour, to the exclusion of the rest of the onset, that is relevant. An interpretation of the initial pitch of the contour as an anchor point for a reference line is inconsistent with our finding that the stimuli with the short contour showed no onset effect. The results rather suggest that a longer stretch of speech is needed in order to estimate the reference line. This is not surprising: the pitch of an initial portion of, say, 50 ms is probably an unreliable basis for making such an estimate. Frequently, such a brief stretch represents the rather variable rising slope up to a pitch peak whose exact beginning will be influenced by such entirely incidental factors as the presence of an unstressed syllable before the peak, the presence of a voiced onset in the accented syllable, or even whether the syllable nucleus consists of a long or a short vowel. The aim of experiment V, which we report in the next section, was to see if this reference line descends as a function of time.

C. Experiment V: Onset length, or the declination effect revisited

The purpose of experiment V was, first, to replicate the ''Onset Height'' effect found in experiment IV, and second, to establish whether there is a difference in prominence between peaks following short onsets and peaks following long onsets. If so, this would support the hypothesis that the prominence of peaks is estimated on the basis of a reference line which declines as a function of time. The results of Pierrehumbert (1979), which showed that the perceived pitch of a later peak is greater than that of an earlier peak, suggest that this is in fact so. However, in that experiment, the timing of the peak was confounded with its serial position, and although Gussenhoven and Rietveld (1988) found a declination effect in one-peak contours, we felt that the declination effect needed to be replicated with one-peak contours in order to establish that it is elapsed time, rather than preceding intonational structure, that causes it. If it is found that later peaks have greater perceived prominence than earlier peaks even after unaccented, level-pitched stretches of speech, then the case for time dependence becomes stronger. In addition, we wanted to test whether the effect of onset height that was found in experiment IV could be replicated. Specifically, we wanted to see if the two effects, the declination effect and the onset effect, could be shown to be independently active in the same perception experiment.

1. Methods

In order to produce the sentence with the short onset while controlling the segmental composition of the accented words, we excised the fragment *de hele dag*, which was exactly 500 ms long, from the speech file containing the sentence *Dat geblaat de hele dag van die schapen daar* (sentence 2 in experiment IV). In this way we obtained two sentences, which served as the basis of our stimuli:

(1) Dat geblaat van die SCHAPen daar (1715 ms) ''that bleating of those SHEEP there''
(2) Dat geblaat de hele dag van die SCHAPen daar (2215 ms) ''that bleating all day of those SHEEP there''

The peaks in the artificial contours were varied in three steps (140, 150, 160 Hz), the onsets in four steps (100, 108, 116, 122 Hz), while the offset was 87 Hz in all stimuli, exactly as in experiment IV. The 24 stimuli were presented to 30 subjects from the same population as before. In other respects, the methods were the same as those in experiments III and IV.

2. Results

Figure 9 gives the mean prominence ratings for the different combinations of onset and peak height for the different sentences. An analysis of variance (repeated measures) was carried out on the scores averaged over the three replications, with three fixed within-subject factors: ''Peak Height,'' ''Onset Height,'' and ''Onset Length;'' the factor ''Subjects'' was regarded as a random factor. All three main effects were significant: ''Peak Height'' [$F(2,58)=63.4$, $p<0.01$], ''Onset Height'' [$F(3,87)=16.1$ $p<0.01$], and ''Onset Length'' [$F=(1,29)=20.2$, $p<0.01$]. Also, the interaction effect between ''Onset Length'' and ''Onset Height'' was significant [$F(3,87)=9.5$, $p<0.01$].

In addition to the effect of ''Peak Height,'' which caused higher peaks to be judged as more prominent, the factor ''Onset Height'' was significant: When the onset was raised, the prominence of the peak decreased. This replicates the result of experiment IV. The factor ''Onset Length,'' too, had the predicted effect: the later peak had greater perceived prominence (61.4 for the short onset versus 63.9 for the long onset). The effects are shown in Fig. 9. The interaction effect between ''Onset Length'' and ''Onset Height'' means that raising the onset did not have the same effect for short and long onsets. Inspection of Fig. 9(a) and (b) suggests that the depressing effect of raised onsets on the perceived prominence of the peak was more pronounced for long onsets than for short onsets: for onset heights 100, 108, 116, and 122 Hz, respectively, mean prominence ratings were obtained of 64.3, 63.3, 58.2, and 60.0 for the short onset and 66.7, 64.6, 64.1, and 60.5 for the long onset.

3. Discussion

The results of experiment V replicated those of experiment IV. First, as we expected, higher peaks are perceived to be more prominent. Second, onset height affects the prominence judgments for an $F0$ peak, for onsets that were longer than 400 ms. We interpret this result as the effect of raising

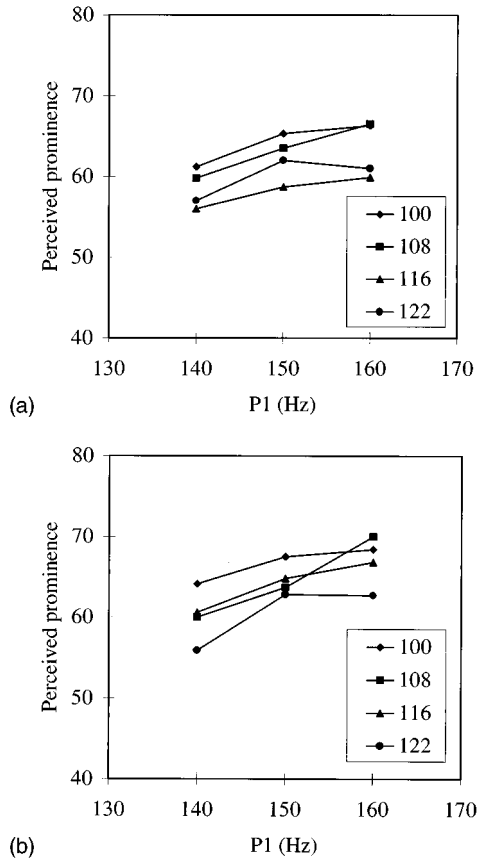


FIG. 9. Mean perceived prominence ratings on a 100-mm scale for *SCHApen* (''sheep'') as a function of Peak Height and Onset Height, for different onset lengths. (a) Short onset (Sentence 1): *dat geblaat van die SCHApen daar*; (b) Long onset (Sentence 2): *dat geblaat de hele dag van die SCHApen daar*.

the reference line that listeners use to estimate the prominence of the peak. This reference line can be determined better as the listener has more information about the onset pitch.

An additional finding is that the length of the onset makes a difference. When the onset is longer, the perceived prominence of the peak is greater: 61.4 for short onsets versus 63.9 for long onsets. The finding that distance from onset is relevant can be interpreted as a replication of the declination effect reported for English by Pierrehumbert (1979). However, in her pitch range experiment, the effect of declination was shown in two-peak stimuli of the same length: The second peak had greater perceived pitch than the first when $F0$ values of the peaks were the same (for the 71-Hz pitch range, cf. Pierrehumbert's Fig. 3). The results of our experiment V show that the declination effect is purely length dependent.

An interaction effect that might have been expected is that between ''Peak Height'' and ''Onset Length.'' In experiments I and II it was found that a higher $P1$ required a larger physical difference between $P1$ and $P2$ in order to obtain equal prominence. From this it can concluded that greater $P1$ prominence is correlated with a more steeply declining abstract reference line. Accordingly, we might expect the difference in perceived prominence between the low peaks in the short stimuli and the low peaks in the long

stimuli to be smaller than that between the high peaks in the short stimuli and the high peaks in the long stimuli. It should be borne in mind, however, that the effect of ''Onset Length'' was itself rather small. Further data are needed to demonstrate an even subtler interaction between onset length and peak height.

The unexpected interaction between ''Onset Length'' and ''Onset Height'' implies that the effect of a short onset is not the same as that of a long onset. While the results are extremely orderly, with prominence increasing with ''Peak Height,'' decreasing with ''Onset Height,'' and increasing with ''Onset Length,'' the detailed results show that the prominence judgments for peaks in the short stimuli with the highest onset (122 Hz) are higher than expected. Possibly, the reversal of the ''Onset Height'' effect in the short stimuli is due to listeners' uncertainty about the category of the onset pitch. In the descriptions of Dutch intonation, a low onset contrasts with a high onset ('t Hart *et al.*, 1990; Gussenhoven, 1988), with a high onset causing the contour to sound more lively and thus likely to evoke greater perceived prominence for the $F0$ peaks.[3] Conceivably, our highest onset in the short sentence was ambiguous between a categorically low and a categorically high onset.

III. GENERAL DISCUSSION AND CONCLUSION

We have reported five experiments that sought to answer the question how listeners determine the prominence of $F0$ peaks in one-peak and two-peak intonation contours. The following main findings were obtained:

(1) Prominence relationships in the female and male pitch ranges are perceived in similar ways (experiments I and II).
(2) For two peaks $P1$ and $P2$, with $P2$ occurring later in the utterance than $P1$, $P2$ should be lower than $P1$ in absolute frequency in order to be perceived as having the same prominence as $P1$ (experiments I and II).
(3) A single peak $P1$ with a fixed frequency is perceived as more prominent if it occurs later in the utterance (experiment V).
(4) A change in the distance between $P1$ and the observable baseline by increasing the rate of baseline declination does not require a concomitant change in the distance between $P2$ and the baseline in order to preserve equal perceived prominence (experiments I and II).
(5) The height of the contour's final low pitch has no effect on the perceived prominence of the peaks in the contour (experiment III).
(6) Increasing the frequency of $P1$ while keeping the baseline fixed requires a small increase in the difference between $P1$ and $P2$ in order to maintain equal prominence; this (small) effect is more pronounced for the female than for the male speaker (experiments I and II).
(7) Increasing the rate of baseline declination (i.e., raising the onset) while keeping $P1$ fixed requires a small decrease in the difference between $P1$ and $P2$ in order to maintain equal prominence; this effect applies both to the female and male speaker, but again it is a small effect (experiments I and II).
(8) Raising the onset for a given $P1$ leads to a reduction of the perceived prominence for $P1$ (experiments IV and V)
(9) The effect of onset height on perceived prominence depends on the amount of onset information. Raising the onset affects the perceived prominence for a peak only with long onsets (over 400 ms) (experiments IV and V).

With these results, we can now address the question of how the perceived prominence of $F0$ peaks relates to characteristics of the $F0$ contour.

The first factor affecting perceived prominence is the position of the $F0$ peak in the utterance: peaks in utterances with long onsets are judged to be more prominent than peaks in utterances with short onsets (finding 3). Likewise, in utterances containing two peaks, the second peak should be lower than the first peak in order to be perceived as having the same prominence (finding 2). These findings provide further evidence for Pierrehumbert's (1979) proposal that listeners expect $F0$ to decline in the course of the utterance as a compensation for actual $F0$ declination. It may be captured by postulating a reference line that declines as a function of time. The fact that actual $F0$ in the stimuli in experiments IV and V did not decline, implies that the reference line is abstract rather than being induced by the observable $F0$ declination rate.

The second factor affecting perceived prominence is related to the size of the $F0$ change underlying the peak. This size can be altered in two ways. One is by altering the height of the peak while keeping its base fixed. The other is by altering the base while keeping the peak fixed. In experiments IV and V, both modifications were found to influence perceived prominence: if the size of the change is reduced by lowering the peak or raising the onset, prominence will decrease; if the size of the change is increased by raising the peak or lowering the onset, prominence will also increase (provided the onset is long enough to give rise to a stable pitch percept; finding 7).

The results of experiments I, II, IV, and V [and also the results in Terken (1991)] suggest that it is not the actual size of the change, measured as the distance between the peak and the observable baseline, which determines the degree of perceived prominence. For instance, in experiments I and II the actual size for $P1$ could be changed rather drastically by changing the slope of the baseline, while only small adjustments in the size of the change underlying $P2$ were needed in order to maintain equal prominence (cf. Figs. 2, 3, and 4). Instead, it would appear that what counts is the distance to the abstract reference line as postulated above, which determines the prominence associated with a peak. This was also the central assumption in the model proposed by Ladd (1993).

In Ladd's (1993) model, the perceived prominence of $P1$ and $P2$ is determined by their distance to the abstract reference line. Ladd assumes that the slope of the abstract reference line is fixed regardless of the contour's position in the speaker's overall pitch range; furthermore, he assumes that equal prominence is obtained when the pitch maxima have equal distance to the reference line.

The latter assumption is contradicted by the current findings in two ways. First, it predicts that raising or lowering the reference line up or down the frequency axis (for instance, by raising or lowering the onset) leaves the peaks equidistant, no matter what metric (Hz, st, or ERB rate) is employed, and that therefore declination rate should not affect the $P1-P2$ difference giving equal prominence. The current results show that there is an effect of rate of baseline declination on the $P1-P2$ difference, even though it is relatively small, so that this prediction of Ladd's model is refuted. Second, it predicts that an increase in the height of $P1$ should cause an equal increase in the height of an equally prominent $P2$: The distance between $P2$ and the reference line ($D2$) should still increase by the same amount as that between $P1$ and the reference line ($D1$) when $P1$ is raised. The results of experiments I and II show that this is not the case. Rather, it appears that, if the distance between the peak and the abstract reference line ($D'1$) is large, the difference between $P1$ and $P2$ needed to maintain the impression of equal prominence is larger than when $D'1$ is small. The finding that the size of the $P1$ change ($D'1$) affects the difference between $P1$ and $P2$ required to obtain equal prominence is not compatible with the hypothesis that the degree of perceived prominence is a direct function of the distance between the peak and the abstract baseline. However, as observed in the discussion of experiment V, these results can quite naturally be interpreted to mean that the slope of the abstract reference line is steeper as the $P1$ prominence is higher. Equivalently, we could assume a fixed slope for the reference line, in combination with an iso-prominence line through $P1$ and $P2$ which declines more steeply as the prominence of $P1$ is higher, in line with Pierrehumbert (1980: Chap. 3).

From these observations the contours of a model for explaining prominence perception can be derived. Perceived prominence appears to be related to the distance between the peak and an abstract reference line, which declines at a rate which is independent of the observable baseline. Increases in the distance between $P1$ and the abstract baseline lead the listener to expect the reference line to descend at a faster rate. We cannot at this point provide quantitative estimates of the different effects; further and more fine-grained experiments should provide the required data.

A further finding which deserves attention is the fact that the effect of onset height was found only in the condition in which the distance between the beginning of the contour and the peak was at least two words long. Apparently, the listener requires a certain amount of speech to be able to make a reliable estimate of the onset pitch. This has two interesting implications. In the first place, it means that the initial one or two syllables apparently do not contain sufficient information to accurately estimate onset height, in contrast with what would be expected on the basis of many current theories. In the second place, it implies that in utterances with short onsets, where the listener apparently is unable to estimate onset height, the abstract baseline is not anchored at all in the actual contour. At present we cannot answer the question how the position of the contour in the speaker's overall range is estimated in these cases; it seems likely that, lacking sufficient information, the listener assumes a default range. This would also apply in cases where there is no onset information at all, such as in the case where the first syllable is accented.

In summary, it appears that perceived prominence of accent peaks is determined mainly by the distance between the peak and an abstract reference line, which declines as a function of time and is anchored in the onset of the utterance (provided that the onset is long enough to enable the listener to obtain a reliable estimate of onset height). In addition, there are clear indications that greater $P1$ prominence is associated with more steeply declining reference lines.

ACKNOWLEDGMENTS

Experiment I was conducted while Bruno Repp was a Visiting Research Fellow at the Institute for Perception Research (IPO). The support of the Technical University Eindhoven and of the University of Nijmegen in the preparation and running of the experiments is gratefully acknowledged. Special thanks are due to Leo Vogten for his cheerful and effective help with various technical problems in the case of experiments I and II and to Rob van den Berg for his contribution to the successful completion of experiments III, IV, and V. We thank the colleagues and students who served as subjects in the rather tedious experiments. We are indebted to Mary Beckman, Bob Ladd, and an anonymous reviewer for helpful comments on draft versions.

[1]Two additional subjects were tested, one in the judgment task only and the other in both tasks. The former almost always judged $P2$ to be more prominent than $P1$; the later produced extremely variable results. Their data were excluded. One of the authors (B. Repp), though not a speaker of Dutch, took part in both tests as a pilot subject and produced results consistent with the subject averages presented below.

[2]In anticipation of this effect, the range of $P2$ values assigned to each $P1$ value had been centered around a value slightly below $P1$. It might be argued that this asymmetry in stimulus design actually caused the effect. This seems unlikely, however, in view of the randomization of all stimuli in the judgment task, and the consistent presence of a $P1-P2$ difference in earlier studies. The same difference was obtained on those trials in the adjustment task where the subjects started with the highest $P2$ value, although it was smaller.

[3]A reviewer has pointed out to us that the reduced prominence on *scha–* in the long onset condition may be due to the listener's interpretation of the preceding *–blaat* as having a H^* pitch accent. Under that view, the prominence of $P2$ might be heard as reduced either because there is an accent-to-accent declination effect, or a phrase-to-phrase declination effect, as opposed to a purely time-dependent declination effect. There are two reasons why this alternative account is improbable. The interpretation of the level onset pitch as containing a H^* pitch accent should be most likely when the onset is highest. However, in the short condition, it is precisely the highest onset that shows a reversal of the reduction of the perceived prominence on $P2$, a finding that cannot be squared with this alternative explanation. Also, the alternative explanation might lead one to expect a more abrupt effect of ''Onset Height'' rather than the gradual effect that is observed.

Beattie, G. W., Cutler, A., and Pearson, M. (**1982**). ''Why is Mrs. Thatcher interrupted so often?,'' Nature (London) **300**, 744–747.
Beckman, M., and Pierrehumbert, J. (**1986**). ''Intonational structure in English and Japanese,'' Phonology Yearbook **3**, 255–310.
Finney, D. J. (**1971**). *Probit Analysis* (Cambridge U.P., New York).
Fujisaki, H., and Hirose, K. (**1984**). ''Analysis of voice fundamental frequency contours for declarative sentences of Japanese,'' J. Acoust. Soc. Jpn. (E) **5**, 233–242.
Geluykens, R., and Swerts, M. (**1994**). ''Prosodic cues to discourse boundaries in experimental dialogues,'' Speech Commun. **15**, 69–77.
Gussenhoven, C. (**1988**). ''Adequacy in intonation analysis: The case of Dutch,'' in *Autosegmental Studies on Pitch Accent*, edited by H. van der Hulst and N. Smith (Dordrecht, Cincinnati), pp. 95–121.
Gussenhoven, C., and Rietveld, A. C. M. (**1988**). ''Fundamental frequency declination in Dutch: testing three hypotheses,'' J. Phonetics **16**, 355–369.
't Hart, J., Collier, R., and Cohen, A. (**1990**). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody* (Cambridge U.P., U.K.).
Hermes, D. J. (**1988**). ''Measurement of pitch by subharmonic summation,'' J. Acoust. Soc. Am. **83**, 257–264.
Ladd, D. R. (**1993**). ''On the theoretical status of ''The Baseline'' in modeling intonation,'' Lang. Speech **36**, 435–451.
Ladd, D. R., and Terken, J. M. B. (**1995**). ''Modelling intra- and inter-speaker pitch range variation,'' in *Proceedings of XIIIth International Congress of Phonetic Sciences, Vol. II*, edited by K. Elenius and P. Branderud (Stockholm University, Stockholm), pp. 386–389.
Liberman, M., and Pierrehumbert, J. (**1984**). ''Intonational invariance under changes in pitch range and length,'' in *Language Sound and Structure*, edited by M. Aronoff and R. Oehrle (MIT, Cambridge, MA), pp. 157–233.
Mozziconacci, S. J. L. (**1995**). ''Pitch variations and emotions in speech,'' in *Proceedings of XIIIth International Congress of Phonetic Sciences*, Vol. I, edited by K. Elenius and P. Branderud (Stockholm University, Stockholm), pp. 178–181.
Pierrehumbert, J. B. (**1979**). ''The perception of fundamental frequency declination,'' J. Acoust. Soc. Am. **66**, 363–369.
Pierrehumbert, J. B. (**1980**). *The Phonology and Phonetics of English Intonation* (MIT, Cambridge, MA).
Rietveld, A. C. M., and Gussenhoven, C. (**1985**). ''On the relation between pitch excursion size and prominence,'' J. Phonetics **13**, 299–308.
Rump, H. H., and Collier, R. (**1996**). ''Focus conditions and the prominence of pitch-accented syllables,'' Lang. Speech **39**, 1–17.
Sluijter, A. M. C. (**1991**). ''Een perceptieve evaluatie van een model voor alinea-intonatie met synthetische spraak'' (A perceptual evaluation of a model for paragraph intonation with synthetic speech). Internal Report 801. Institute for Perception Research, Eindhoven, The Netherlands.
Sluijter, A. M. C., and Terken, J. (**1993**). ''Beyond sentence prosody: Paragraph intonation in Dutch,'' Phonetica **50**, 180–188.
Swerts, M., Collier, R., and Terken, J. (**1994**). ''Prosodic predictors of discourse finality in spontaneous monologues,'' Speech Commun. **15**, 79–90.
Terken, J. (**1991**). ''Fundamental frequency and perceived prominence of accented syllables,'' J. Acoust. Soc. Am. **89**, 1768–1776.
Terken, J. (**1994**). ''Fundamental frequency and perceived prominence of accented syllables II: Nonfinal syllables,'' J. Acoust. Soc. Am. **95**, 3662–3665.

Study of woodwind-like systems through nonlinear differential equations. Part I. Simple geometry

Ana Barjau
Departament d'Enginyeria Mecànica, Universitat Politècnica de Catalunya, Diagonal 647, 08028 Barcelona, Spain

Vincent Gibiat and Noël Grand
Laboratoire Ondes et Acoustique, ESPCI, URA CNRS 1503 Université Paris 7 Denis Diderot, 10 rue Vauquelin, 75231 Paris Cedex 05, France

(Received 7 September 1996; revised 13 February 1997; accepted 24 June 1997)

Idealized woodwind models assume a well-localized nonlinearity coupled to a linear bore whose mathematical description is usually expressed through a convolution product. For the case of the simplest bore geometries, cylindrical and conical, the convolution can be transformed into a delayed differential equation. When including in it the nonlinearity, the usual concepts of nonlinear dynamical systems allow a better understanding of the system's evolution. In this paper they are applied to the cylindrical and conical geometries and some characteristics of their behaviors are analyzed at the threshold of oscillation through an analytical study. It is shown that their dynamics cannot be reduced to a finite number of degrees of freedom (effective dynamics). Finally numerical simulations reveal peculiar characteristics of the direct and inverse bifurcations involved in such simple systems for quadratic and cubic nonlinearities respectively. © *1997 Acoustical Society of America.* [S0001-4966(97)01410-0]

PACS numbers: 43.75.Ef [WJS]

INTRODUCTION

Traditionally, woodwinds have been represented through idealized models which assume a linear part (the bore) and a point localized nonlinear part (the reed and more generally the excitatory system) at the bore entrance section. Their mutual interaction leading to the self-sustained oscillatory behavior is represented through the diagram shown in Fig. 1. Their time-domain study has been based on algebraic or differential equations for the nonlinear part and on an integral or convolution equation coupling the reed to the linear part:

$$\frac{du(t)}{dt} = F_{\mathrm{NL}}((u(t),p(t),P_0), \tag{1a}$$

$$p(t) = h(t) * u(t), \tag{1b}$$

where F_{NL} is a nonlinear algebraic equation and $h(t)$ is the impulse response at the bore entrance section. This formulation can be very suitable for simulation, but does not give any evident clues to find the values of the parameters leading to all possible behaviors of the system. Particularly, it is not easy to predict under which conditions stable and unstable oscillations will take place and which degrees of freedom govern the amplitude of the oscillation at steady state.

Some integral studies have been carried out (Worman, 1971; Schumacher, 1978; Grand, *et al.*, 1994c), but they often need heavy and complicated calculations partially made in the frequency domain to obtain final results in the time domain. It is not surprising since the frequency domain is not well suited for nonlinear systems.

The study of the stability of a nonlinear dynamical system (NLDS) can be done with the classical tools and concepts of this area provided the system description is made entirely through differential equations. Even if the dynamics of chaos is currently a field of increasing interest, the basic concepts, tools, and experiments date from the nineteenth century. It is not our purpose to give extensive information about this domain, which the reader can easily find in references (Bergé *et al.*, 1984; Lauterborn and Parlitz, 1988). The aim of the following work is to show that for a cylindrical and a conical resonator, Eqs. (1a) and (1b) can be transformed into nonlinear delayed differential equations. Since these geometries are far from real instruments, a second paper is devoted to the case of a real geometry (Barjau and Gibiat, 1997) which is dealt with using a different approach.

In order to transform the convolution equation (1b) into a differential one, it is necessary to have an analytical expression for the kernel $h(t)$. Theoretical results are available for the cylinder (Polack *et al.*, 1987), which accurately fit the experimental data. For the geometries with which we are concerned, an *ad hoc* formulation is also possible which correctly reproduces Polack's results. The nonlinear delayed differential equations derived from this *ad hoc* formulation can then be studied through the concepts of NLDS. The system bifurcations are analyzed from the system simulation with the help of a fast numerical algorithm (Barjau *et al.*, 1990).

I. *AD HOC* FORMULATION FOR THE TIME RESPONSE OF THE CYLINDER AND THE CONE

Various models can be found in the literature for the simplest geometry, the cylinder. If the radiation at the open end and the viscothermal losses are neglected, $h(t)$ reduces to a comb of alternate Dirac impulses (Maganza, 1985). Another model has been used by McIntyre *et al.* (McIntyre *et al.*, 1983) where a Gaussian-type reflection function $r(t)$

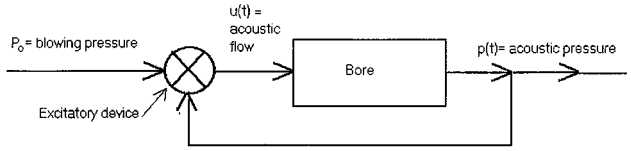


FIG. 1. Schematic diagram of an idealized woodwind. The feedback loop represents the mutual interaction between the linear and the nonlinear parts of the system.

is used instead of $h(t)$. Both models do not respect the basic asymmetry of time-domain responses. More recently Polack *et al.* (Polack *et al.*, 1987) have developed the analytical calculation for the reflection function of a cylindrical pipe. Their result is mathematically complicated but in very good agreement with the experimental data.

The use of the reflection function $r(t)$ instead of $h(t)$ leads to a shorter time-domain description of the cylindrical bore, greatly reducing the time required to compute the convolution equation, which is now written as

$$p(t)-\frac{Z_0}{S}u(t)=r(t)*\left[p(t)+\frac{Z_0}{S}u(t)\right], \tag{2}$$

and can be seen as a simple change of variables (McIntyre *et al.*, 1983; Maganza, 1985), where Z_0 is the characteristic impedance of the air and S is the bore entrance section area. For other shapes, $r(t)$ is not necessarily shorter but it is always lower than $h(t)$ (Schumacher, 1981).

It is always advisable to work with the minimal response (the shortest possible one) for every possible bore geometry. The reflection function $r(t)$ is the minimal response only for the cylindrical case. For the conical case the minimal response is provided by the use of the so-called modified impulse response $h_{\mathrm{mod}}(t)$ (Agulló *et al.*, 1988). This concept is equivalent to $r(t)$ for the cylindrical case, the only difference being a Dirac $\delta(t)$ distribution and a multiplying constant. The shift from $h(t)$ to $h_{\mathrm{mod}}(t)$ can be seen from a mathematical point of view as a change of variables:

old variables: $(p(t),u(t))$

old equation: $p(t)=h(t)*u(t)$;

new variables: $(p(t),n(t))$

new equation: $p(t)=h_{\mathrm{mod}}(t)*n(t)$.

Both the cone and the cylinder share the same $h_{\mathrm{mod}}(t)$, although the function $n(t)$ to be convolved with it to obtain $p(t)$ is different. For the cylindrical bore,

$$p(t)=h_{\mathrm{mod}}(t)*n(t)=h_{\mathrm{mod}}(t)*\left[u(t)+\frac{S}{Z_0}p(t)\right], \tag{3}$$

and for the conical one,

$$p(t)=h_{\mathrm{mod}}(t)*\left[u(t)+\frac{S}{Z_0}p(t)+\frac{S}{Z_0}\frac{c}{r}\int_0^t p(t')dt'\right], \tag{4}$$

where r is the distance from the entrance section of the truncated cone to the full cone apex.

We proposed a formulation for $h_{\mathrm{mod}}(t)$ (Grand and co-workers, 1994a, b, c) based on an exponential function

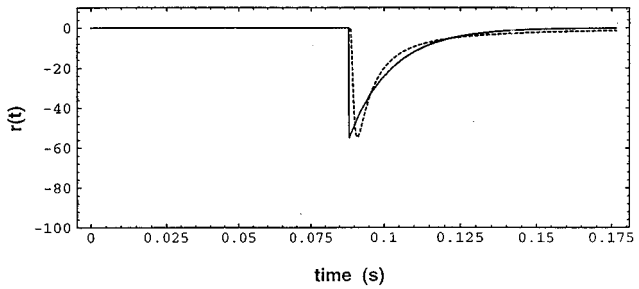


FIG. 2. Comparison between our approach (continuous line) and Polack's formulation (dotted line) for the case of a cylindrical tube of length 15 m and diameter 5 mm, $\sigma=70\ \mathrm{s}^{-1}$, $\beta=55$.

which allows the transformation of the convolution into a differential equation while remaining close to the exact solution given by Polack *et al.* (Polack *et al.*, 1987):

$$h_{\mathrm{mod}}(t)=\frac{1}{2}\frac{Z_0}{S}\left[\delta(t)-\beta\epsilon(t-T)\exp(-\sigma(t-T))\right], \tag{5}$$

where $\delta(t)$ and $\epsilon(t)$ are the Dirac and the Heavyside distributions, respectively, T is the time needed for a round trip in the bore, and σ and β are related to the losses and properly adjusted to fit the experimental data.

The choice of an exponential function is not surprising. On the one hand, its mathematical properties (easy derivation and integration) simplify and make possible the analytical manipulations. On the other hand, it is often used to describe the time-extended decay of variables in physical phenomena with damping (as for example the string in a plucked string; Cremer, 1984).

It is clear that this description corresponds neither to the physical frequency-dependent damping due to viscothermal losses, nor to the radiation condition at the open end of the tube usually found in the literature (Bruneau, 1983), but it is a sufficiently accurate description for our purpose and it is an improvement compared to the previous models since it respects the basic asymmetry of the real reflection function. For the case of a cylindrical tube of length 15 m and diameter 5 mm, Fig. 2 shows the comparison between our approach and Polack's formulation with $\beta=55\ \mathrm{s}^{-1}$ and $\sigma=70\ \mathrm{s}^{-1}$. These dimensions have been chosen because they correspond to the example studied theoretically and experimentally by Polack *et al.* They do not correspond to a musical instrument, but the purpose of this article is not the study of a musical instrument—are there musical instruments of absolutely cylindrical or conical shape?—but to work with general systems consisting on a nonlinearity coupled to a duct and to obtain hints on the possible behaviors of musical instruments.

II. NONLINEAR DELAYED DIFFERENTIAL EQUATIONS

The Laplace transform of Eqs. (3) and (4) combined with Eq. (5) yields Eqs. (6) and (7) for the cylinder and the cone, respectively (Grand and co-workers, 1994a, b, c):

$$\frac{dp(t)}{dt}+\sigma p(t)+\beta p(t-T)$$
$$=\frac{Z_0}{S}\left[\frac{du(t)}{dt}+\sigma u(t)-\beta u(t-T)\right], \tag{6}$$

$$\frac{d^2p(t)}{dt^2}+\left[\sigma-\frac{c}{r}\right]\frac{dp(t)}{dt}-\sigma\frac{c}{r}p(t)$$
$$+\beta\frac{dp(t-T)}{dt}+\beta\frac{c}{r}p(t-T)$$
$$=\frac{Z_0}{S}\left[\frac{d^2u(t)}{dt^2}+\sigma\frac{du(t)}{dt}-\beta\frac{du(t-T)}{dt}\right]. \tag{7}$$

Actually, Eq. (6) can be obtained from Eq. (7) by taking $r\to\infty$.

These equations are general and apply for any description of the excitatory system as the coefficients β and σ can take any value (according to the chosen tube) and the nonlinear function $u(t)$ has not been specified yet. The only important assumption (or restriction) is that of the one-point spatial localization of the nonlinearity. The main characteristic of these two equations is that they involve delayed values of the variables. In other words, their integration at any time t calls for the knowledge of an infinite set of initial conditions (corresponding to the time evolution of the variables over the previous time interval of length T). So, from a physical point of view, they describe an infinite-degree-of-freedom system which can in principle exhibit any kind of dynamic behavior, from the simple classical ones (static, periodic...) to more complicated ones (quasi-periodic, chaotic,...).

However, from an intuitive point of view, it is clear that if T is small compared to the characteristic times of the physical system (relaxation, damping,...) the number of degrees of freedom can be dramatically reduced. At the limit of $T\to 0$ Eqs. (6) and (7) would imply just one and two degrees-of-freedom, respectively.

Nonlinear differential equations with delay are not as rare in the nonlinear dynamical systems domain, particularly when studying natural self-oscillating phenomena; to give just a few examples: leukocyte concentration in human blood (Mackey and Glass, 1977), and evolution of animal population in closed ecological systems (May, 1980). They also appear in numerous physical systems: laser (Le Berre *et al.*, 1986), hydrodynamic systems (Villermaux, 1993), etc.

If T is not negligible, it is well-known that the system behavior at steady state is governed by a reduced number of degrees-of-freedom, or modes (Manneville, 1991). According to a classical classification we will talk of ''central modes'' and ''marginal'' or ''slave modes,'' the central modes being those whose damping becomes zero at the threshold of oscillation (thus undergoing a bifurcation) and which govern the amplitude of the marginal ones through a nonlinear relation at steady state (see the Appendix). This reduced description is not always possible, but when it exists, it is known as the system effective dynamics. Of course such a description is meaningless at the transient state, so numerical simulations of steady state are unavoidable to validate the simplification. If the effective dynamical description is not possible, the system oscillations depend on all its modes in principle and then the steady-state amplitude will be limited by the nonlinearity of the system.

We will first consider the simplest possible (quadratic) nonlinear function coupling the flow and the acoustic pressure at the bore entrance section (McIntyre *et al.*, 1983):

$$u(t)=\begin{cases}\alpha[P_0-p(t)][p(t)-p_f], & \text{for } p(t)>p_f,\\ 0, & \text{for } p(t)\leqslant p_f,\end{cases} \tag{8}$$

where P_0 is the blowing pressure, p_f the pressure under which no flow is introduced in the system, and α is a properly adjusted coefficient. A cubic function will be used later on to show the different behavior to which it gives rise.

As we will see in the next section, the lack of description for the reed dynamics does not invalidate the method, as such dynamics can always be introduced through a finite number of additional modes.

III. SYSTEMATIC STUDY OF MARGINAL AND CENTRAL MODES

In order to obtain the effective dynamical description, it is necessary to determine the central or bifurcating modes. They depend in principle on the values of the system parameters. In our study the blowing pressure P_0 (or any function of it) is a natural control parameter since it is one of those that is controlled by the musician and the only one easily measurable.

The systematic study of the effective dynamics begins with the determination of the equilibrium or fixed points, and is followed by a linear analysis at their neighborhood and a search for the bifurcating modes.

A. Cylindrical case

The delayed differential equation (6) together with the definition given for the air flow in Eq. (8) leads to

$$A\left[\frac{dp(t)}{dt}+\sigma p(t)\right]=-\alpha\frac{Z_0}{S}\left[\frac{dp(t)}{dt}+\sigma p(t)\right]p(t)$$
$$+\beta\left[A-2+\alpha\frac{Z_0}{S}p(t-T)\right]$$
$$\times p(t-T)+B, \tag{9}$$

where the parameters A and B are functions of P_0 with $A=1-\alpha(Z_0/S)(P_0+p_f)$ and $B=\alpha(Z_0/S)(\beta-\sigma)P_0p_f$. The equation defining the two equilibrium positions, that is, the values of the internal pressure corresponding to nonoscillatory solutions, $dp/dt=0$, is

$$A(P_0)\sigma p_{\rm eq}=-\alpha\frac{Z_0}{S}\sigma p_{\rm eq}^2+\beta\left[A(P_0)-2\right.$$
$$\left.+\alpha\frac{Z_0}{S}p_{\rm eq}\right]p_{\rm eq}+B(P_0).$$

The stability of such positions is given by the sign of the time derivative of the pressure:

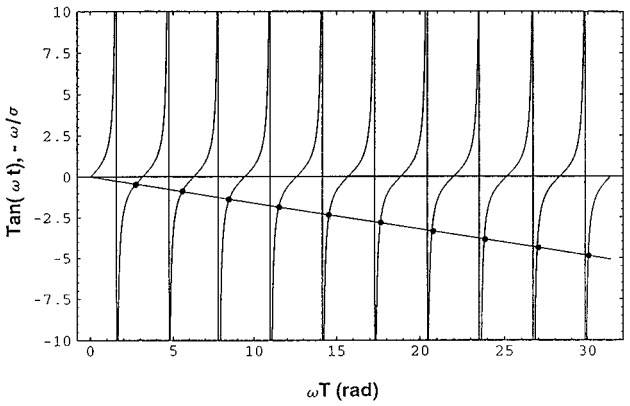


FIG. 3. Graphical solution of the transcendental Eq. (13). The intersection (dots) between $\tan(\omega T)$ and ω/σ (straight line) give the bifurcating modes for the cylinder of Fig. 2.

$$\left.\frac{dp(t)}{dt}\right|_{\text{eq}} = \frac{\beta[\lambda(P_0)-2]-\sigma\lambda(P_0)}{\lambda(P_0)}\, p_{\text{eq}} = (\beta-\sigma)-\frac{2\beta}{\lambda(P_0)}. \tag{10}$$

The linear approximation at the neighborhood of p_{eq} [pressure $= p_{\text{eq}} + p(t)$ with $p(t)$ small], leads to

$$\lambda(P_0)\,\frac{dp(t)}{dt} + \sigma\lambda(P_0)p(t) = \beta[\lambda(P_0)-2]p(t-T), \tag{11}$$

with $\lambda(P_0) = A(P_0) + 2\alpha(Z_0/S)p_{\text{eq}}$ as the control parameter since it is P_0 dependent.

Substitution of a solution of the form $p(t) = p(0)\exp((\xi + j\omega)t)$ in Eq. (11) splits it into two equations (real and imaginary parts):

$$\lambda(\xi+\sigma) = \beta(\lambda-2)\exp(-\xi T)\cos(\omega T), \tag{12a}$$

$$\lambda\omega = -\beta(\lambda-2)\exp(-\xi T)\sin(\omega T). \tag{12b}$$

The bifurcating modes will be those for which $\xi = 0$, and are given by the transcendental equation

$$\tan(\omega T) = -\frac{1}{\sigma T}\,(\omega T). \tag{13}$$

This equation is graphically solved simply by studying the intersection of the two functions defined by its left- and right-hand sides, as presented in Fig. 3 (σ and T correspond to the cylinder of Fig. 2; even if that case is not a musical one it allows a clear visualization).

One important feature of Eq. (13) is that losses appear only on the right-hand side. It is the slope of the straight line which carries this information. For negligible losses ($\sigma \to \infty$) the straight line becomes horizontal and the solutions of Eq. (13) form a complete harmonic set and correspond both to the ''resonances'' *and* to the ''anti-resonances'' of the amplitude impedance curve of the cylinder with an open end. This result points in the same direction as that stated by Grand (Grand, 1994c). As losses grow, the ''resonance'' solutions tend to form a set of odd harmonic values. If losses are infinite ($\sigma = 0$), we have just the solution $\omega = 0$, which means that no oscillation will take place. Except for the latter case, the number of bifurcating modes is always infinite (although countable) and independent of the control parameter, and the reduction of the model to an effective one is not possible.

This clearly indicates that a perfect cylinder, with a perfectly localized and instantaneous nonlinear part, is, at its threshold of oscillation, a perfectly unstable system where *all* the modes can participate. When σ is infinite, the modes are harmonic so one can expect a periodic solution. When σ is not infinite, the modes are no longer harmonic and the transient will contain growing nonharmonic modes. The periodicity, if any, is then obtained through a phase locking process governed by the nonlinearity at higher amplitude.

Another interesting result which can be obtained straight away from Eqs. (12a) and (12b) is that of the existence of a nonoscillatory mode with damping independent of the control parameter. By setting $\omega \to 0$, we obtain

$$\left.\begin{aligned}\lambda(\xi_0+\sigma) &= \beta(\lambda-2)\exp(-\xi_0 T)\\ \frac{\lambda}{T} &= -\beta(\lambda-2)\exp(-\xi_0 T)\end{aligned}\right\} \Rightarrow \xi_0 = -\left(\frac{1}{T}+\sigma\right). \tag{14}$$

In order to know what is actually going to happen when we ''blow'' into the cylindrical tube it is necessary to perform numerical simulations: The nonoscillatory mode could lead to a fixed point (equilibrium point) or be ''killed'' by the bifurcating ones. The possibility of a limit cycle cannot be studied analytically as the equilibrium points. Such behaviors can only be determined by numerical simulation. If a limit cycle is found, it is necessarily a stable one, as unstable behaviors are rarely detectable.

B. Conical case

The delayed differential equation (7) together with Eq. (8) yields

$$\begin{aligned}&\frac{d^2p(t)}{dt^2} + \left(\sigma - \frac{c}{r}\right)\frac{dp(t)}{dt} - \sigma\,\frac{c}{r}\,p(t)\\ &\quad + \beta\left[\frac{dp(t-T)}{dt} + \frac{c}{r}\,p(t-T)\right]\\ &\quad = \alpha\,\frac{Z_0}{S}\left\{[p_f + P_0 - 2p(t)]\left[\frac{d^2p(t)}{dt^2} + \sigma\,\frac{dp(t)}{dt}\right]\right.\\ &\qquad \left. - 2\left(\frac{dp(t)}{dt}\right)^2 - \beta[p_f + P_0 - 2p(t-T)]\,\frac{dp(t-T)}{dt}\right\},\end{aligned} \tag{15}$$

and setting the derivatives equal to zero, we obtain the equation giving the only possible equilibrium position:

$$(\beta-\sigma)\,\frac{c}{r}\,p_{\text{eq}} = 0. \tag{16}$$

The linear approximation at the neighborhood of $p_{\text{eq}} = 0$ is

$$(1-A)\frac{d^2p(t)}{dt^2}+\left[\sigma(1-A)-\frac{c}{r}\right]\frac{dp(t)}{dt}-\sigma\frac{c}{r}p(t)$$
$$=-\beta\left[(1+A)\frac{dp(t-T)}{dt}+\frac{c}{r}p(t-T)\right], \tag{17}$$

with $A=\alpha(Z_0/S)(p_f+P_0)$ as the only control parameter. The substitution of a solution of the form $p(t)=p(0)e^{(\xi+j\omega)t}$ in Eq. (17) gives two equations (real and imaginary parts):

$$(1-A)(\xi^2-\omega^2+\sigma\xi)-\frac{c}{r}(\xi+\sigma)$$
$$=-\beta\exp(-\xi T)\left\{\left[\xi(1+A)+\frac{c}{r}\right]\cos(\omega T)+\omega(1+A)\sin(\omega T)\right\}, \tag{18a}$$

$$(1-A)(2\xi\omega+\sigma)-\frac{c}{r}\omega$$
$$=-\beta\exp(-\xi T)\left\{\omega(1+A)\cos(\omega T)-\left[\xi(1+A)+\frac{c}{r}\right]\sin(\omega T)\right\}, \tag{18b}$$

and we obtain the bifurcating modes by setting $\xi=0$:

$$\tan(\omega T)=-\omega\frac{\omega^2(1-A^2)-(c/r)(1-2\sigma)}{\omega^2\sigma(1-A^2)-(c/r)(c/r+2\omega^2)}$$
$$=F(\omega), \tag{19}$$

whose graphical solution is given in Fig. 4(a). For a medium value of the losses, the solutions form a complete family of quasi-harmonic frequencies. As in the case of the cylinder, the reduction to an effective model is impossible.

The nonoscillatory modes are defined by a cubic equation obtained by setting $\omega=0$ in Eqs. (18a) and (18b):

$$(A^2-1)\xi^3+\left[(A^2-1)(1+\sigma)+2\frac{c}{r}\right]\xi^2+\frac{c}{r}\left[2(A+\sigma+1)-\frac{c}{r}\right]\xi+\frac{c}{r}\left[\frac{c}{r}(1-\sigma)-2\sigma\right]=0. \tag{20}$$

Solution of Eq. (20) yields three roots whose dependence on the control parameter P_0 is shown in Fig. 4(b). Actually, two of them remain negative for the calculated range of blowing pressures (and so correspond to stable modes), while the third one soon becomes positive, giving rise to an unstable mode which can help the bifurcating ones in establishing an oscillation. As in the cylindrical case, the zero amplitude threshold oscillation will be necessarily multimodal with nonharmonic modes, and a periodic steady state will only be possible if the nonlinearity succeeds in the phase locking of those modes.

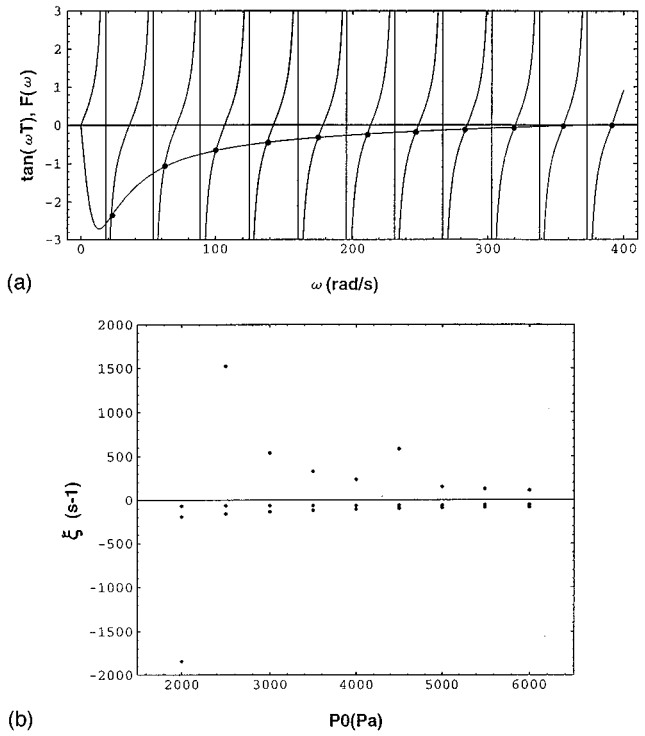


FIG. 4. (a) Graphical solution of the transcendental Eq. (19) giving the bifurcating modes for a cone of length 15 m, with $\sigma=70\,s^{-1}$. Entrance diameter is 4 mm and final diameter 64 mm. (b) Damping of the nonoscillating modes for the same cone.

IV. TIME-DOMAIN SIMULATION OF CYLINDRICAL AND TRUNCATED CONICAL ''WOODWINDS''

As stated in Sec. III A, the actual result of ''blowing'' into the instrument can only be studied through the numerical integration of Eqs. (9) and (15). As both geometries, cylindrical and conical, are assumed to be described by a time response of the exponential type in this paper, such an integration can be reduced to an iterative algorithm of the form (Barjau *et al.*, 1990):

$$p(t+\Delta t)=a_1p(t)+a_2n(t)+b_1p(t-\Delta t)+b_2n(t-\Delta t)+c_1p(t-T)+c_2n(t-T)+d_1p(t-\Delta t-T)+d_2n(t-\Delta t-T), \tag{21}$$

where $a_{1,2}$, $b_{1,2}$, $c_{1,2}$, $d_{1,2}$ are constant coefficients and $n(t)$ is the function defined in Sec. I.

The physical information that can be obtained through simulation is not only the value P_0 at the threshold of oscillation (that is, the value at which a first bifurcation appears), but also the type of bifurcation that the system undergoes—direct or inverse.

It has been shown that this type of bifurcation is related to the particular shape of the nonlinear function chosen to represent the excitatory system. A quadratic function implies an inverse bifurcation—discontinuous jump from zero to finite amplitude oscillation—while a cubic one produces a direct bifurcation—continuous transition from zero to finite amplitude of oscillation. This has been noted first by McIntyre *et al.* (McIntyre *et al.*, 1983), although described with different words, explained by Maganza *et al.* (Maganza

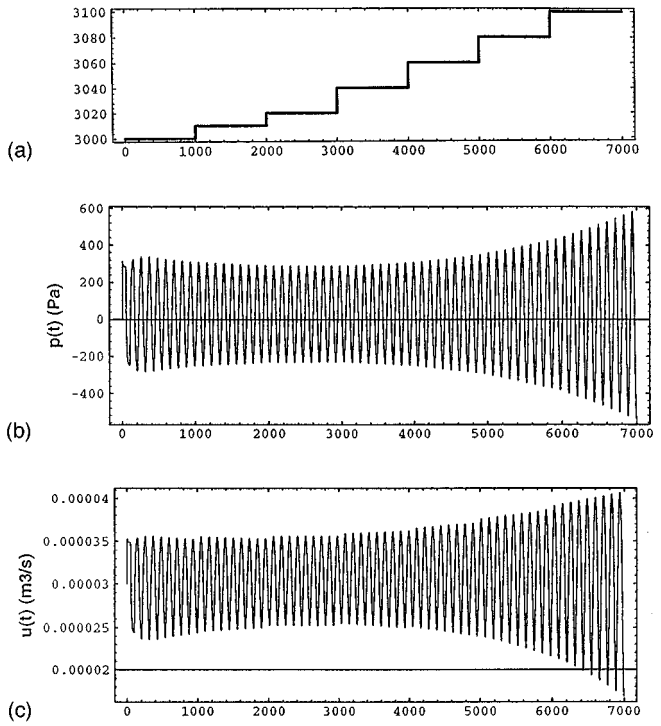


FIG. 5. Pressure (b) and flow (c) obtained with P_0 values (a) increasing from 3000 Pa to 3100 Pa and with a quadratic nonlinearity for the cylinder of Fig. 2.

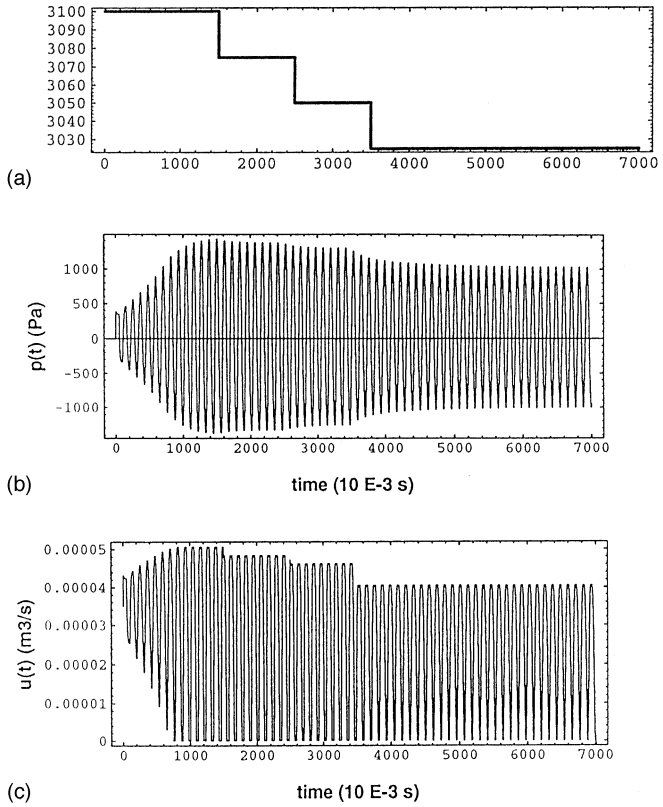


FIG. 6. Pressure (b) and flow (c) obtained with P_0 values (a) decreasing from 3100 Pa to 3000 Pa and with a quadratic nonlinearity for the cylinder of Fig. 2.

et al., 1987) and proved by Grand (Grand *et al.*, 1994c). If the variation of the control parameter leads to a direct bifurcation, a stable oscillation with very low infinitesimal amplitude appears after the threshold. If the bifurcation taking place is an inverse one, the resulting oscillation after threshold has a finite amplitude. Another interesting difference is the existence of an hysteretic phenomenon when the control parameter is decreased for an inverse bifurcation: The oscillation can be maintained with a value of the control parameter lower than the threshold one.

Since a musical instrument is supposed to be played between pianissimo and fortissimo levels, some former works have examined for a direct bifurcation in woodwinds. For the case of the cylinder defined in Fig. 2 and a quadratic nonlinear function, Fig. 5(b) and (c) shows the pressure $p(t)$ and air flow $u(t)$ at the entrance section, respectively, the results for an interval of P_0 values going from 3000 Pa to 3100 Pa [Fig. 5(a)]. The threshold of oscillation is located at about 3050 Pa.

Figure 6(b) and (c) shows $p(t)$ and $u(t)$ at the same section but decreasing values of P_0 going from 3100 Pa to 3000 Pa [Fig. 6(a)]. The graphics clearly show a hysteretical behavior of the system, as a stable oscillation can be attained at $P_0 = 3025$ Pa. That is, the bifurcation leading to a limit cycle from an equilibrium position is an inverse one and the width of the P_0 zone where the hysteretical behavior takes place, ΔP_0, is about 30 Pa (1% of P_0). Figure 7(a) and (b) shows the corresponding bifurcation diagram. The low value of ΔP_0 [Fig. 7(b)] indicates however that the bifurcation is not far from being a direct one.

If the nonlinearity is a cubic one,

$$u(t) = \alpha(P_0 - p(t))(p(t) - p_f)(P_0 + p(t) - 2p_f),$$

the bifurcation becomes a direct one. Simulations for increasing and decreasing values of P_0 have been carried out and do not reveal any hysteresis. Figure 8(a) and (b) presents

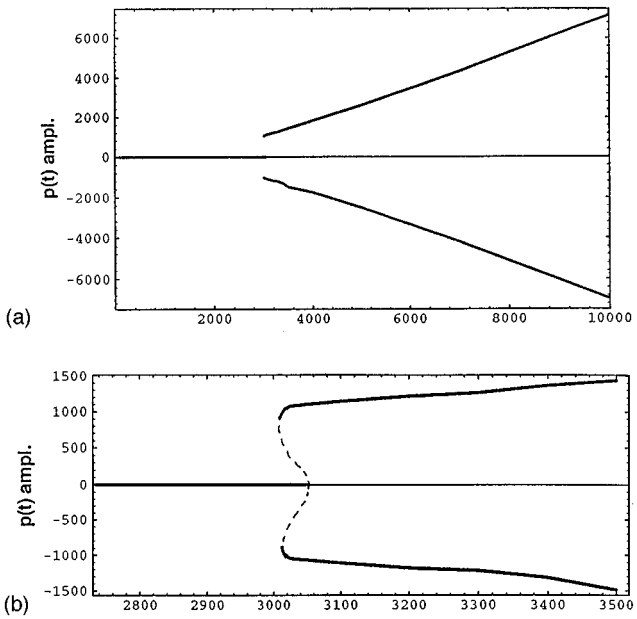


FIG. 7. Inverse bifurcation diagram for a range of P_0 values going from 0 Pa to 10 000 Pa (a) and from 2750 Pa to 3500 Pa (b) for the cylinder of Fig. 2 and the quadratic nonlinearity (the dotted line represents the unstable branches).

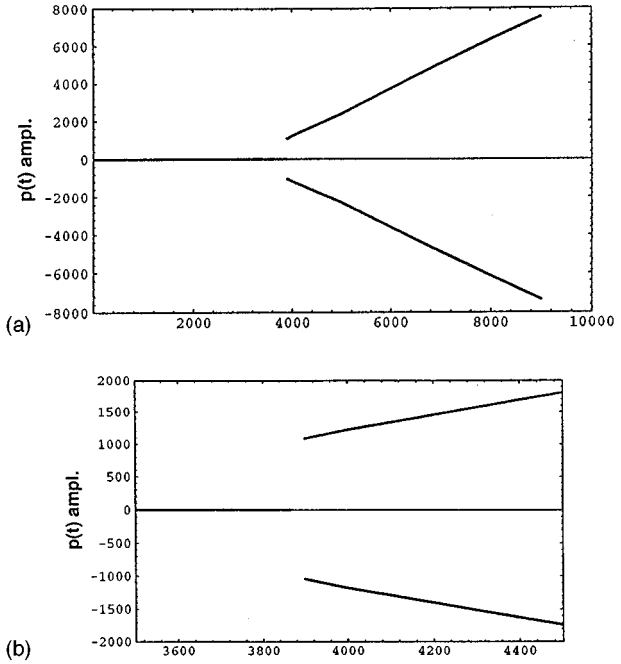


FIG. 8. Direct bifurcation diagram for a range of P_0 values going from 0 Pa to 10 000 Pa (a) and from 3500 Pa to 4500 Pa (b) for the cylinder of Fig. 2 and the cubic nonlinearity.

the bifurcation diagram in this case. Nevertheless, comparison of Figs. 7(a) and 8(a) makes it evident that the difference, even if interesting and certainly from a physical point of view, is perfectly negligible from a ''musical'' point of view.

For the case of the truncated cone, the stability of the simulation is much more dependent on the choice of the parameters. For the case of the quadratic nonlinearity, the threshold of oscillation has been found to be located in the neighborhood of $P_0=5200$ Pa. Figure 9 shows the first oscillations for the smooth conicity case of Fig. 4 and a blowing pressure of 5300 Pa. Other simulations have been performed for a wide range of blowing pressures, and their results are schematized in the bifurcation diagram in Fig. 10. It is a direct bifurcation.

Another collection of simulations has been calculated for a different value of the α parameter in the flow function. It has been observed that, after periodic behaviors, the signal becomes biperiodic (second bifurcation), as it can be easily seen in Fig. 11(a) (beating frequency) or in the two other

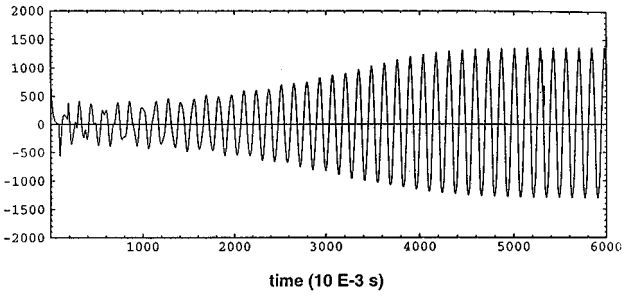


FIG. 9. Pressure obtained with a P_0 of 5400 Pa for the cone of Fig. 4 and the quadratic nonlinearity.

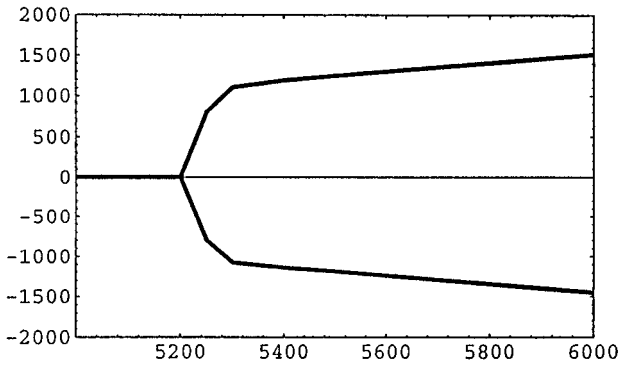


FIG. 10. Bifurcation diagram as a function of the blowing pressure for the cone of Fig. 4 and the quadratic nonlinearity.

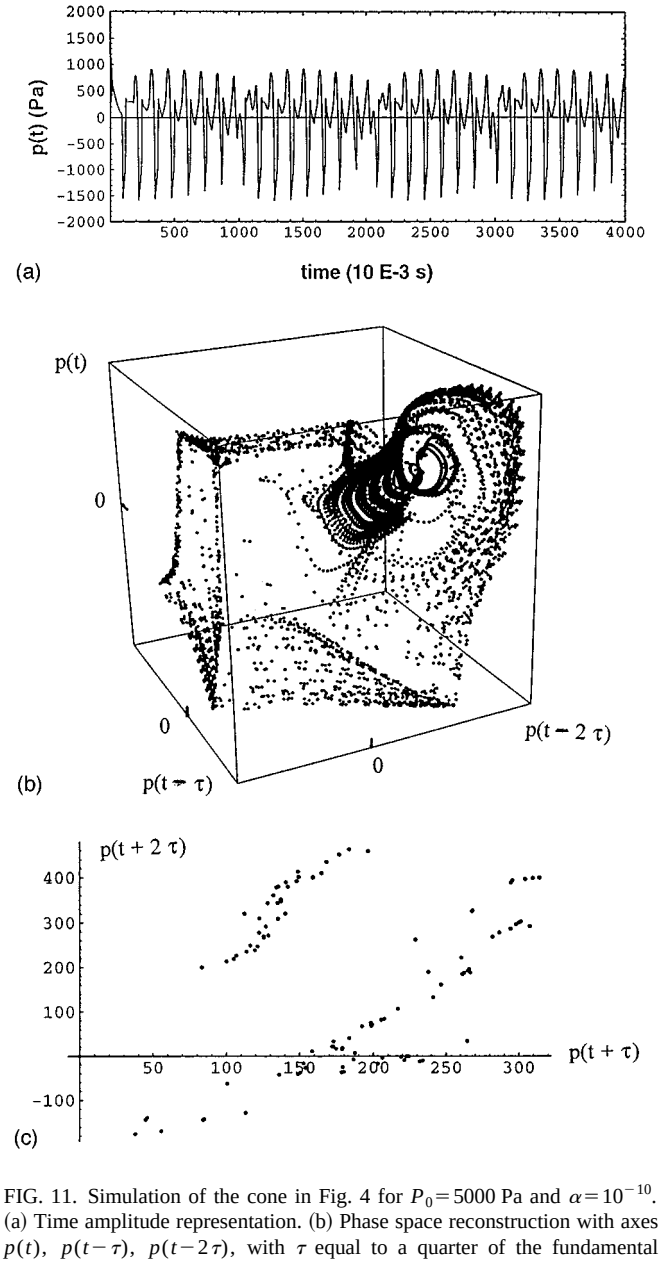


FIG. 11. Simulation of the cone in Fig. 4 for $P_0=5000$ Pa and $\alpha=10^{-10}$. (a) Time amplitude representation. (b) Phase space reconstruction with axes $p(t)$, $p(t-\tau)$, $p(t-2\tau)$, with τ equal to a quarter of the fundamental period of the oscillation. (c) Poincaré section by the plane $p(t)=0$ of the phase space reconstruction.

representations, phase space reconstruction [toruslike shape, Fig. 11(b)] and Poincaré section [Fig. 11(c)]. This result is not surprising since the delayed differential equation for the cone contains three different characteristic times. The first one is the delay associated with the round trip of the acoustical wave. The second one is the characteristic time of the damping, very small compared to the former and which does not seem to be relevant in the biperiodic state. The third one is the inverse of the eigenfrequency appearing in the nondelayed part of the left-hand side of Eq. (17):

$$\frac{\sigma c}{r(1-A)}.$$

Changes in the conicity (changes in r) affect this characteristic time and thus the signal, although still biperiodic, will show a different beating frequency.

As conicity increases, the simulation becomes more and more sensitive to small changes in the parameters, and the accumulation of numerical truncation errors results in diverging nonphysical behaviors which can only be avoided by dramatically reducing the sampling interval, increasing the numerical precision, and thus highly increasing the required computation time.

V. CONCLUSION

An analytical formulation of the time responses of strictly cylindrical and conical woodwinds has been presented which leads to a totally differential formulation, whose main characteristic is that it includes delayed variables. Such a description is quite new in the domain of musical acoustics even if frequent in other scientific domains.

The study of stability of the system through a linear approach leads to transcendental equations which show the impossibility of reducing the description of such systems to an effective dynamical one. However, they give important information concerning the modes which play an important role at the threshold of oscillation. For the cylindrical case, the even modes participate with a strength comparable to that of odd ones. This result is obtained in a simpler way in the presented time approach than in previous frequency-domain methods (Grand, 1994c). For the conical tube, all modes (even and odd) participate in the oscillating regime. This means that the threshold of oscillation of perfectly conical or cylindrical systems involves all of their resonant modes. Consequently, since truncated conical bores present inharmonic modes, their transient state at any blowing pressure will contain nonharmonic components, as is easily seen in Fig. 9.

A numerical integration of a cylindrical and a conical system has been carried out for different values of the blowing pressure and for two different nonlinear functions (representing the excitatory system coupled to the bore): quadratic and cubic ones. Whenever the bore is a cylindrical one, the results show that a periodic oscillatory regime is attained through an inverse bifurcation in the quadratic case, while the bifurcation becomes direct if the quadratic nonlinearity is replaced by the cubic one. However, the difference between the direct and the inverse bifurcations is so small that it is, in practice, hardly noticeable.

For the case of the studied conical bore, the bifurcation is direct for the quadratic flow. In the parameter space of the system, different control parameters other than the blowing pressure can be used. If the α parameter appearing in the flow function is decreased, a second bifurcation leading to quasi-periodic behavior appears.

ACKNOWLEDGMENTS

The present work would not have been possible without the financial support of the French ''Ministère des Affaires Etrangères'' and the Spanish ''Ministerio de Educación y Ciencia'' through a co-operation program (AIP 93142, AI175-B). We would like to thank R. T. Schumacher for his helpful comments during the reviewing process.

APPENDIX

For a general unidimensional physical system described through a nonlinear partial differential equation of the type:

$$a\,\frac{\partial^i \Phi(x,t)}{\partial t^i} + b\,\frac{\partial^j \Phi(x,t)}{\partial x^j} = F_{\mathrm{NL}}\left(\frac{\partial^k \Phi(x,t)}{\partial t^k}, \frac{\partial^l \Phi(x,t)}{\partial x^l}\right), \tag{A1}$$

and given initial and boundary conditions, a modal study at the threshold can be done by linearizing the differential equation and trying a separable solution of the form:

$$\Phi(x,t) = T(t)X(x). \tag{A2}$$

The boundary conditions lead to restrictions on $X(x)$, so that $\Phi(x,t)$ becomes a superposition of an infinite number of countable solutions:

$$\Phi(x,t) = \sum_{n=1}^{\infty} T_n(t)X_n(x) \tag{A3}$$

[with well-defined $X_n(x)$]. The time-varying amplitudes $T_n(t)$ are governed by an infinite number of ordinary nonlinear differential equations which can always be expressed, through a suitable change of variables, as first-order ones:

$$\frac{dA_n(t)}{dt} + \sigma_n A_n(t) = F_{\mathrm{NL}}(A_1(t), A_2(t), \ldots, A_m(t)). \tag{A4}$$

A solution to these equations is given by Manneville (1990):

$$A_n(t) = A_n^0 e^{-\sigma_n t} + e^{-\sigma_n t} * F_{\mathrm{NL}}(A_1(t), A_2(t), \ldots, A_m(t)). \tag{A5}$$

It is clear that for $\sigma_n > 0$, both terms become negligible for increasing values of t, that is, far from the transient state there are no changes in $A_n(t)$, and Eq. (A4) becomes an algebraic one:

$$A_n(t) = \frac{1}{\sigma_n} F_{\mathrm{NL}}((A_1(t), A_2(t), \ldots, A_m(t))). \tag{A6}$$

Agulló, J., Barjau, A., and Martínez, J. (**1988**). ''Alternatives to the impulse response $h(t)$ to describe the acoustical behavior of conical ducts,'' J. Acoust. Soc. Am. **84**, 1606–1612.

Barjau, A., Cardona, S., and Jordi, L. (**1990**). ''Fast simulation of the internal pressure in a double-reed woodwind,'' Coll. Mod. Phys. (Grenoble) **13**, 161–178.

Barjau, A., and Gibiat, V. (**1997**). ''Study of woodwind-like systems through nonlinear differential equations. Part II. Real geometry,'' J. Acoust. Soc. Am. **102**, 3032–3037.

Bergé, P., Pomeau, Y., and Vidal, Ch. (**1984**). *L'ordre dans le chaos, vers une approche déterministe de la turbulence* (Hermann, Paris).

Bruneau, M. (**1983**). *Introduction aux théories de l'acoustique* (Université du Maine, Le Mans), Chap. 5 for example.

Cremer, L. (**1984**). *The Physics of the Violin* (MIT, Cambridge, MA), p. 92.

Grand, N., Barjau, A., Gibiat, V., and Derode, A. (**1994a**). ''Instruments à vent: Formulation intégrale ou différentielle?,'' J. Phys. IV **C5**, 621–624.

Grand, N., Barjau, A., and Gibiat, V. (**1994b**). ''Differential formulation a powerful tool for woodwind instrument numerical simulation,'' J. Acoust. Soc. Am. **95**, 2860.

Grand, N. (**1994c**). ''Etude du seuil d'oscillation des systèmes acoustiques non-linéaires de type instrument à vent,'' Doctoral thesis, Université Paris VII.

Lauterborn, W., and Parlitz, U. (**1988**). ''Methods of chaos physics and their application to acoustics,'' J. Acoust. Soc. Am. **84**, 1975–1992.

Le Berre, M., Ressaye, E., Tallet, A., and Gibbs, H. M. (**1986**). ''High dimensional attractor of a nonlinear ring cavity,'' Phys. Rev. Lett. **56**, 274–277.

Mackey, M. C., and Glass, L. (**1977**). ''Oscillation and chaos in physiological control systems,'' Science **197**, 287–289.

Maganza, C. (**1985**). ''Excitations non-linéaires d'un conduit acoustique cylindrique. Observations de doublements de période précedant un comportement chaotique. Application à la clarinette,'' Doctoral thesis, Université Du Maine Le Mans.

Maganza, C., Caussé, R., and Laloë, F. (**1986**). ''Bifurcations, period doublings and chaos in clarinet like systems,'' Europhys. Lett. **1**, 295–299.

Manneville, P. (**1990**). ''Dissipative Structures and weak turbulence'' (Academic, New York), see particularly Chap. 5

May, R. M. (**1980**). ''Nonlinear phenomena in ecology and epidemiology,'' Ann. (N.Y.) Acad. Sci. **357**, 267–281.

McIntyre, M., Schumacher, R., and Woodhouse, J. (**1983**). ''On the oscillations of musical instruments,'' J. Acoust. Soc. Am. **74**, 1325–1345.

Polack, J., Meynial, X., Kergomard, J., Cosnard, C., and Bruneau, M. (**1987**). ''Reflection function of a plane sound wave in a cylindrical tube,'' Rev. Phys. Appl. **22**, 331–337.

Schumacher, R. T. (**1978**). ''Self-sustained oscillations of the clarinet: An integral equation approach,'' Acustica **40**, 298–309.

Schumacher, R. T. (**1981**). ''*Ab initio* calculations of the oscillations of the clarinet,'' Acustica **48**, 72–85.

Villermaux, E. (**1993**). ''Auto-oscillation et mélange dans les écoulements recirculants,'' Doctoral thesis, Université Pierre et Marie Curie, Paris VI.

Worman, W. E. (**1971**). ''Self-sustained non-linear oscillations of medium amplitude in clarinetlike systems,'' Ph.D. dissertation, Dept. of Physics, Case Western Reserve University, Cleveland, OH.

Study of woodwind-like systems through nonlinear differential equations. Part II. Real geometry

Ana Barjau
Departament d'Enginyeria Mecànica, Universitat Politècnica de Catalunya, Diagonal 647, 08028 Barcelona, Spain

Vincent Gibiat
Laboratoire Ondes et Acoustique, ESPCI, URA CNRS 1503 Université Paris 7 Denis Diderot, 10 rue Vauquelin, 75231 Paris Cedex 05, France

(Received 7 September 1996; revised 24 June 1997; accepted 24 June 1997)

In a previous paper [Barjau *et al.*, J. Acoust. Soc. Am. **102**, 3023–3031 (1997)], a differential formulation has been used to study the stability and bifurcations in geometrically simple woodwindlike systems. For general bore geometries, such differential formulation can be obtained by representing the system through a lumped equivalent mechanical model whose parameters are extracted from a real impedance curve on which a modal analysis has been performed. The present procedure is then applied to a particular case and several time-domain simulations are carried out in order to both assess the validity of the systematically developed procedure and explore the possible system behaviors far from the threshold of oscillation. © *1997 Acoustical Society of America.* [S0001-4966(97)01510-5]

PACS numbers: 43.75.Ef [WJS]

INTRODUCTION

In a previous paper (Barjau *et al.*, 1997), a simple and systematic study of stability and bifurcations in geometrically simple woodwindlike systems was proposed through analysis of the differential equations derived from their integral form description. As the cylindrical and conical geometries studied in the aforementioned article are far from representing real bore geometries, we propose here a general approach which allows a wholly differential description leading to a straightforward analysis of the instrument stability as a function of a control parameter. In this paper, the excitatory system will be modeled in the same way as in Barjau *et al.* (1997) (that is, through an algebraic nonlinear polynomial equation relating the air input flow to the internal pressure at the bore entrance section) and only the bore model will be considered.

Since the bore is considered a totally linear subsystem, a lumped equivalent mechanical model, whose behavior is described through differential equations, can be established through a simple modal analysis (Agulló and Barjau, 1986; Barjau, 1987). It is then obvious that the dynamics of the excitatory system, the reed, could be easily introduced as a complementary mode.

Of course, this mechanical model is an approximate one. As far as the bore behavior is concerned, the approximation neglects the higher resonant modes. Such a simplification is justified by an impedance measurement: the significant peaks in it are always reduced in number. The lack of any differential equation describing the reed dynamics can be justified by the fact that its resonance frequency lies above the instrument working frequency, and thus its inertia can be neglected, at least for our purpose.

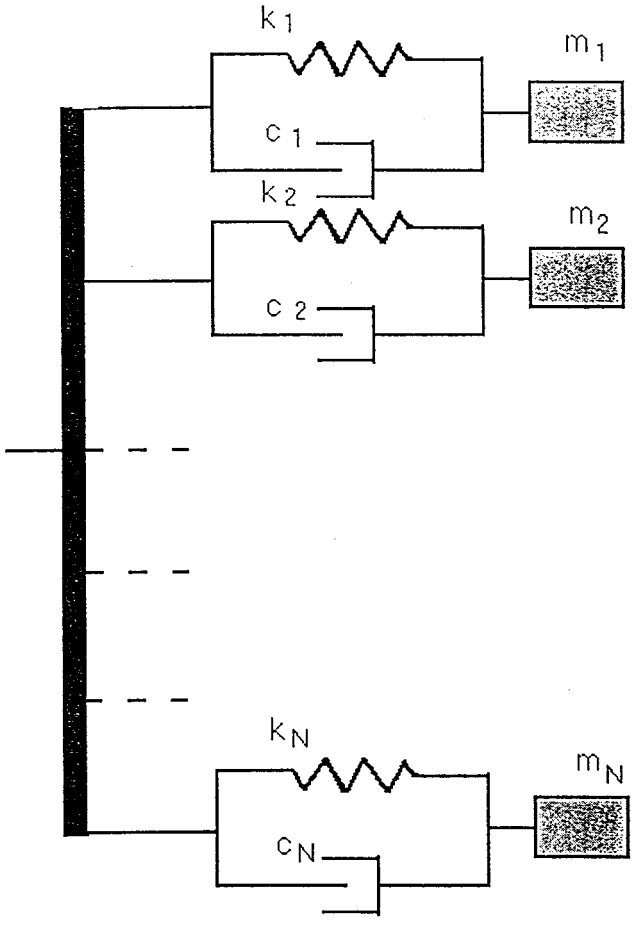


FIG. 1. Lumped equivalent mechanical system for a general bore geometry.

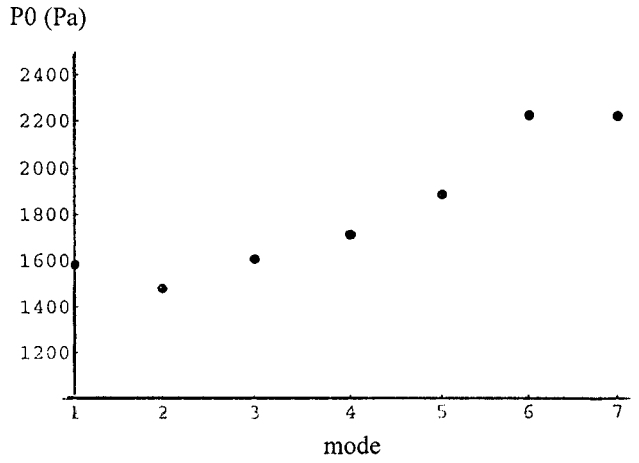


FIG. 2. P_0 values for which the ''tenora'' modes bifurcate (thresholds of oscillation).

I. MODAL DESCRIPTION OF THE IMPULSE RESPONSE

For real geometries, an experimental measurement of the input impedance $Z(\omega)$ is the starting point for the construction of the lumped equivalent mechanical system. Each important resonance of $Z(\omega)$ is associated with a mass-stiffness-damping elementary system whose impulse response will be written as $h_n(t)$. The whole bore is then represented by these systems coupled as shown in Fig. 1, and its corresponding impulse response $h(t)$ is given by the sum of all the elementary responses $h_n(t)$:

$$h(t)=\sum_n h_n(t)=\sum_n [c_n\delta(t)+k_n(1-4\zeta_n^2) \times e^{-\zeta_n\omega_n t}\cos(\omega_n\nu_n t+\Phi_n)], \quad (1)$$

with

$$\omega_n=\sqrt{\frac{k_n}{m_n}}, \quad \zeta_n=\frac{c_n}{2m_n\omega_n}, \quad \nu_n=\sqrt{1-\zeta_n^2},$$

and

$$\Phi_n=A\tan\left[\frac{\zeta_n}{\nu_n}\frac{3-4\zeta_n^2}{1-4\zeta_n^2}\right].$$

The values of the modal parameters $(\zeta_n,\omega_n,\Phi_n)$ or (k_n,m_n,c_n) are deduced from the impedance curve (Agulló and Barjau, 1986).

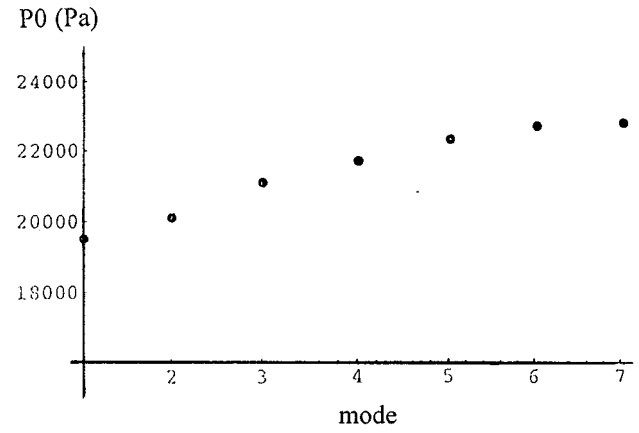


FIG. 3. P_0 values over which the ''tenora'' modes become nonoscillating.

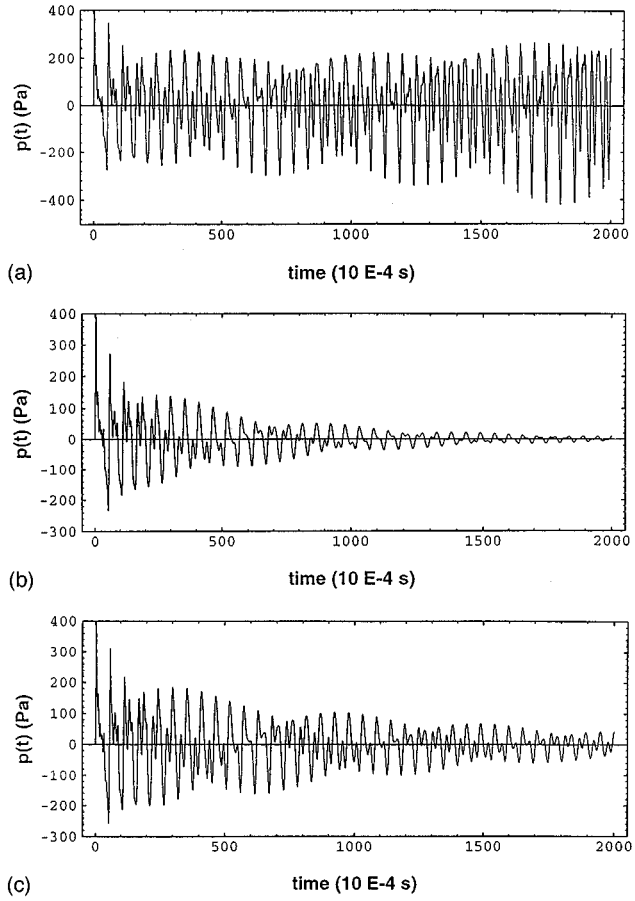


FIG. 4. Time-domain simulations of a ''tenora'' for; (a) $P_0=1482$ Pa; (b) $P_0=1300$ Pa; (c) $P_0=1400$ Pa.

II. DIFFERENTIAL MATRIX FORMULATION

With the modal description, the convolution equation relating the pressure $p(t)$ and the acoustic air flow $u(t)$ at the bore entrance section (S),

$$p(t)=\frac{1}{S}h(t)*u(t)=\frac{1}{S}\sum h_n(t)*u(t)=\sum p_n(t), \quad (2)$$

is easily transformed into a set of ordinary differential equations through a Laplace transform:

$$m_n\frac{d^2p_n(t)}{dt^2}+c_n\frac{dp_n(t)}{dt}+k_np_n(t) =\frac{m_n}{S}\left[c_n\frac{d^2u(t)}{dt^2}+k_n\frac{du(t)}{dt}\right]. \quad (3)$$

For the case of N elementary systems, Eqs. (3) show that the only equilibrium position is $p_n=0$.

The equation describing the air flow $u(t)$ as a function of the internal pressure $p(t)$ and the parameter P_0 (blowing

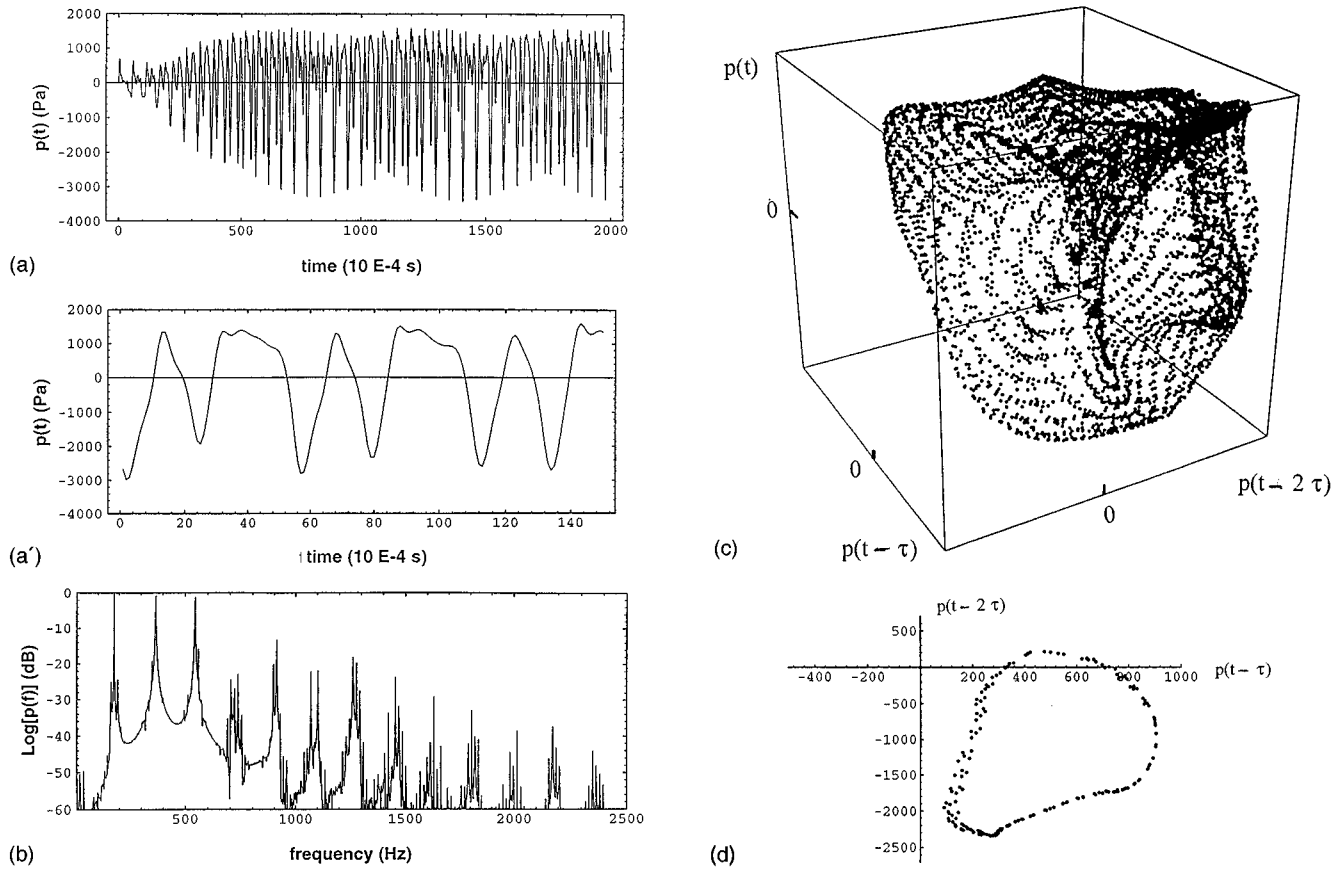


FIG. 5. Simulation of a ''tenora'' for $P_0=2000$ Pa. (a,a′) Time amplitude representation; (b) frequency-domain representation; (c) phase space reconstruction with axes $p(t)$, $p(t-\tau)$, $p(t-2\tau)$; (d) Poincaré section of the phase space reconstruction by the plane $p(t)=0$.

pressure) represents the reed effect on the system. As in (Barjau *et al.*, 1997), a simple quadratic nonlinear model is considered:

$$u(t)=\begin{cases}\alpha[P_0-p(t)][p(t)-p_f], & \text{for } p(t)>p_f,\\ 0 \quad \text{for } p(t)\leqslant p_f,\end{cases}$$

where P_0 is the blowing pressure, p_f the pressure under which no flow is introduced in the system, and α is a properly adjusted coefficient (McIntyre *et al.*, 1983). A more complicated model (taking into account the reed dynamics) would result in adding one or more modes to the modal description.

The linear approach in the neighborhood of the equilibrium position $p_n=0$, once the substitution of the nonlinear relation $u(t)=F_{\mathrm{NL}}[p(t),P_0]$ has been done, gives

$$m_n\frac{d^2p_n(t)}{dt^2}+c_n\frac{dp_n(t)}{dt}+k_np_n(t)$$
$$=\alpha\frac{m_n}{S}\left\{c_n(P_0+p_f)\frac{d^2p(t)}{dt^2}\right.$$
$$\left.+[k_n(P_0+p_f)-2c_n]\frac{dp(t)}{dt}\right\}, \tag{4}$$

which can be rewritten as $2N$ first-order equations:

$$\begin{Bmatrix}\ddot{q}(t)\\ \dot{q}(t)\end{Bmatrix}=[D]\begin{Bmatrix}\dot{q}(t)\\ q(t)\end{Bmatrix}=\begin{bmatrix}M^{-1}C & M^{-1}K\\ I & 0\end{bmatrix}\begin{Bmatrix}\dot{q}(t)\\ q(t)\end{Bmatrix},$$

$$\{q(t)\}^T=\{p_1(t),\ldots,p_N(t)\}, \tag{5}$$

where $M_{ij}=\delta_{ij}-c_i(\alpha/S)(P_0+p_f)$, $C_{ij}=-\delta_{ij}(c_i/m_i)-k_i(\alpha/S)(P_0+p_f)$, $K_{ij}=-\delta_{ij}(k_i/m_i)$, δ_{ij} is the Kronecker symbol, and I is the identity matrix.

III. SYSTEMATIC STUDY OF MARGINAL AND CENTRAL MODES AT THE THRESHOLD OF OSCILLATION

As Eqs. (5) represent a set of coupled differential equations, the $p_n(t)$ do not accept a solution of the harmonic type. However, there is a change of variables,

$$\{p_1(t),\ldots,p_N(t)\}\rightarrow\{\psi_1(t),\ldots,\psi_N(t)\},$$

which diagonalizes the system matrix $[D]$. The new variables accept an harmonic solution whose exponents are nothing but the eigenvalues $\lambda_n(P_0)$ of $[D]$: $\psi_n(t)=\psi_n(0)\exp[\lambda_n(P_0)t]$. The study of their real and imaginary parts, $\xi(P_0)=\mathrm{Re}[\lambda_n(P_0)]$ and $\omega(P_0)=\mathrm{Im}[\lambda_n(P_0)]$, will give all the information about the stability of the fixed point (or equilibrium point) as a function of the control parameter P_0. The bifurcating or central modes, responsible for the instrument behavior at the threshold of oscillation, will be those for which damping $\xi(P_0)$ vanishes. The ones for which damping is negative correspond to marginal modes whose amplitude in an effective dynamical description (only valid in the steady state) is related to that of the central modes through an algebraic nonlinear equation (Manneville, 1990).

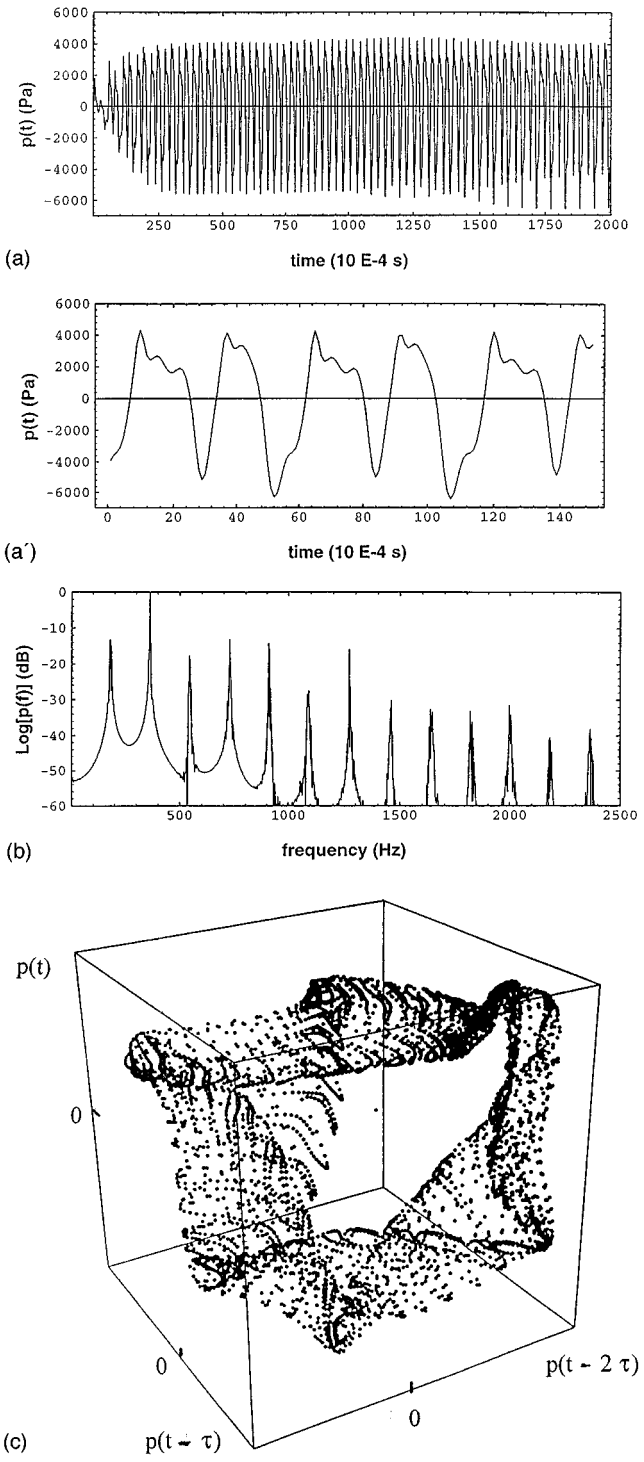


FIG. 6. Simulation of a ''tenora'' for $P_0=4000$ Pa. (a,a′) Time amplitude representation; (b) frequency-domain representation; (c) phase space reconstruction with axes $p(t)$, $p(t-\tau)$, $p(t-2\tau)$.

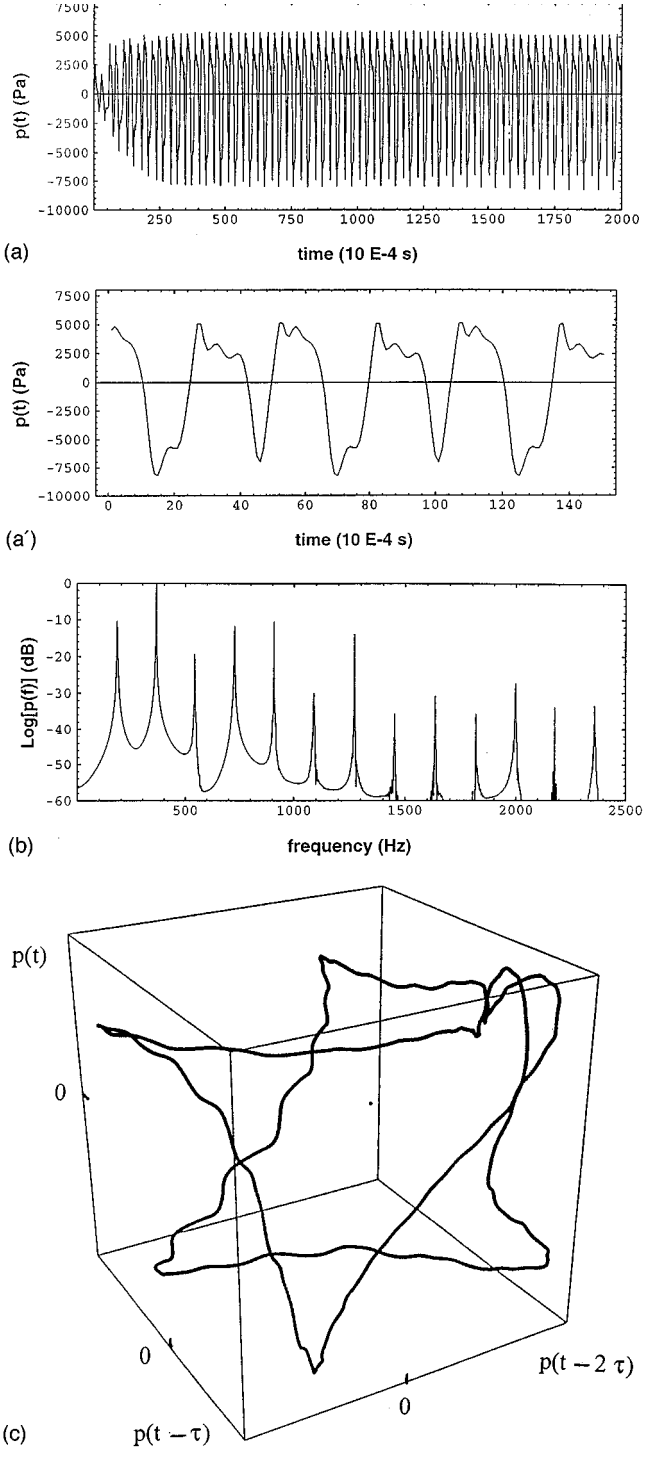


FIG. 7. Simulation of a ''tenora'' for $P_0=5000$ Pa. (a,a′) Time amplitude representation; (b) frequency-domain representation; (c) phase space reconstruction with axes $p(t)$, $p(t-\tau)$, $p(t-2\tau)$.

IV. THE EXAMPLE OF A ''TENORA'' AS A MULTIMODAL SYSTEM

The modal approach has been used to study the behavior of the lowest note (all holes closed) of a ''tenora,'' a double-reed Catalan traditional folk woodwind with a mainly conical bore finished by a bell. The bore has been represented by seven elementary systems whose parameters have been derived from previous measurements (Agulló and Barjau, 1986; Barjau, 1987).

Figure 2 shows the P_0 values of the blowing pressure for which the oscillating modes bifurcate (zero damping). The lowest value from which an unstable behavior can take place is 1482 Pa. Over 2200 Pa the behavior is necessarily unstable as all the modes have bifurcated. The nonlinearity can then succeed in establishing an oscillating steady state of any type (periodic or not). However, one has to keep in mind that such conclusions arise from a linear approach, and that the nonlinearity can give rise to slightly different behaviors.

On the other hand, Fig. 3 gives the P_0 value for which the modes would become nonoscillating. Such behavior can only arise for very high values of P_0 (over 19 000 Pa), never attained in real playing conditions. However, as those values are far from being close to the threshold, the only valid conclusion is the nonexistence of nonoscillating modes in playing conditions.

Several time-domain simulations have been carried out in order to verify the previous conclusions. The particular form of $h_n(t)$ allows the use of a ''fast-convolution'' algorithm (Barjau *et al.*, 1990). We present a selection of these simulations for different values of P_0. It is clear from Fig. 4(a) that at the threshold (1482 Pa) the bifurcating mode succeeds in establishing the oscillation. For two lower pressures Fig. 4(b) and (c) shows the dependence of the signal decay rate upon the deviation from the threshold value.

For higher values of the blowing pressure well above threshold, $P_0=2000$ Pa and $P_0=4000$ Pa, the results are shown in Figs. 5(a) and 6(a), respectively. The steady oscillation seems clear, but it is not so evident to judge if it is a periodic one or not and if the asymptotic regime has been reached. These figures show just the first 20 ms and the last 20 ms of a 10-s computation.

As the blowing pressure increases, the signal gets closer and closer to periodicity. When we go up over 5000 Pa, Figs. 7(a) and 8(a) offer no doubt about periodic signals with a very important second harmonic [easily identified in the zooms in Figs. 7(a′) and 8(a′)].

A traditional tool when analyzing the periodicity of a signal is its Fourier transform (FT). In order to have different views of the phenomena we are dealing with, the FT of the previous time signals [Figs. 5(a)–8(a)] has been done, and the results are shown in Figs. 5(b)–8(b). For the cases $P_0=2000$ Pa and 4000 Pa, we have classical biperiodic (or quasi-periodic with two basis frequencies) multiphonics spectra (Gibiat, 1988). Even if a first glimpse would suggest a periodic structure of about 5.5 ms, a more detailed observation reveals a small and continuous change in the waveform which is not strictly speaking repetitive. This pseudo periodicity is clearly seen in the zoom of Figs. 5(a) and 6(a) [Figs. 5(a′) and 6(a′)]. Maybe a long time computation would lead to a repetitive signal with a very long period. This would be the case for a biperiodic signal with a frequency ratio p/q where p and q are not small integers. However, experimentally speaking and numerically speaking, there are limiting factors (experimental noise, numerical noise, frequency sampling due to time limited DFT, etc.) which would not allow detection of whether the signal is periodic or not.

Because of that uncertainty, another representation, usual in a nonlinear dynamical systems context, has been used: the delayed phase space reconstruction. As usual, such reconstruction is obtained by representing the points in a 3-D space (the minimal dimension allowing chaotic behaviors) defined by $p(t)$ and two delayed values $p(t-\tau)$, $p(t-2\tau)$. If the trajectories in the 3-D space do not intersect, the 3-D model is sufficient.

Even if the choice of the delay τ is crucial when calculating the dimension of the attractor or the Lyapunov exponents, when leading with biperiodic signals the changes in the value τ have the only effect of showing one perspective or another of the picture; in other words, they only change the viewpoint.

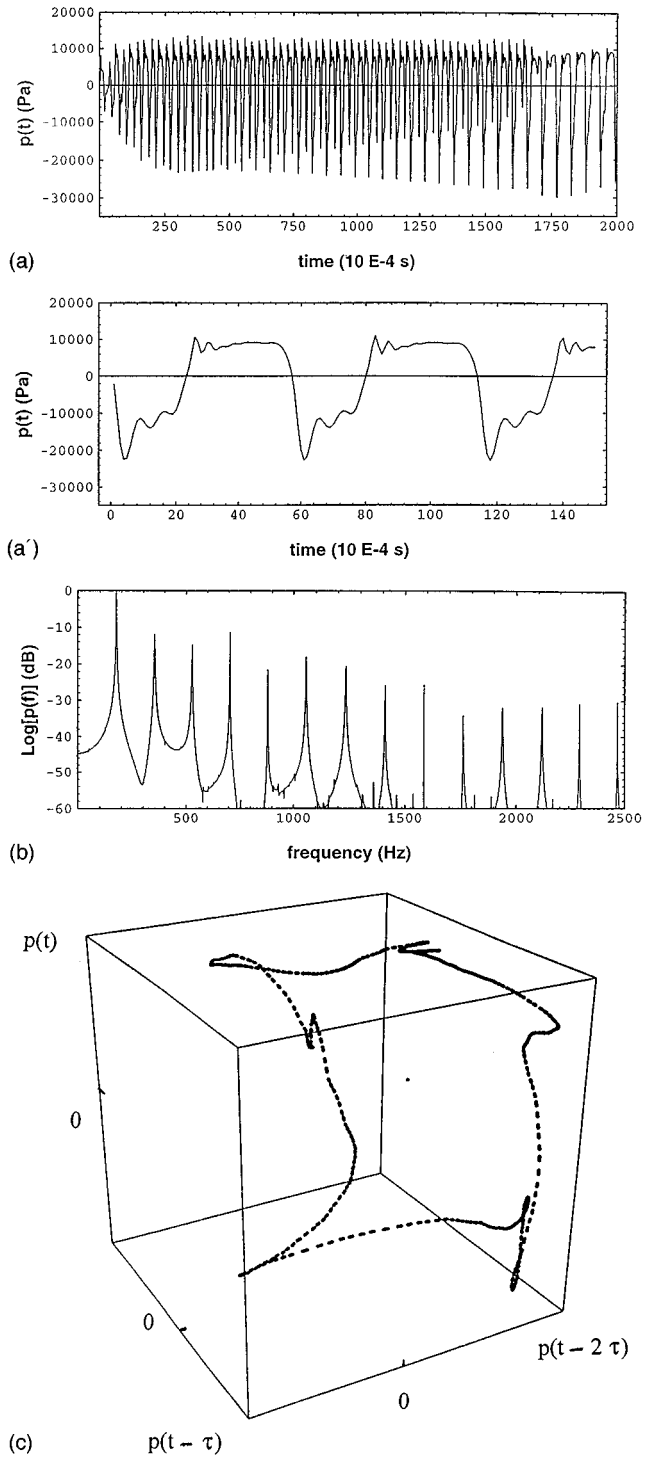


FIG. 8. Simulation of a ''tenora'' for $P_0=10\,000$ Pa. (a,a′) Time amplitude representation; (b) frequency-domain representation; (c) phase space reconstruction with axes $p(t)$, $p(t-\tau)$, $p(t-2\tau)$.

In order to correctly class the behavior in the steady state, the phase space reconstruction has been done using just the last 50 ms of a 2 s duration signal. The results are shown in Figs. 5(c)–8(c). For the case $P_0=2000$ Pa and 4000 Pa, the phase space representation has a toruslike shape (which

would be a dense one if we carried on the simulation for a very long time), the typical one for a biperiodic signal. To verify the assumption of quasi-periodicity suggested by the Fourier transforms of the signal the Poincaré section is one of the best tools. Figure 5(d) shows a closed, but sampled, curve without accumulation points. That means that the signal can be considered a quasi-periodic one with two nonharmonically related basis frequencies (p and q are not small integers). For all the other cases, the continuous line appearing in the phase space reconstruction tells us that the periodic state has been reached. Of course the simulations leading to such figures have been carried out for very long times and very small integration time steps.

It is a common assumption that at least at the threshold of oscillation woodwindlike systems exhibit a sinusoidal behavior (Benade, 1976; Worman, 1971) and that an increase of the blowing pressure results in a growth in the sound harmonic content. Recently Grand (Grand, 1994) has demonstrated that the sinusoidal assumption is only valid for zero amplitude signal (no oscillation!) and has to be replaced by the less general periodic one, that is, that for very low values of the blowing pressure, the signal may vary from a quasi-sinusoidal shape to a squared one.

Our results for the studied example point in an even more surprising direction: The signal starts quasiperiodic to end up in a periodic steady state only when a high value of the blowing pressure has been attained. However, this result is familiar to the ''tenora'' players since they know very well that the lowest note does not accept a *pianissimo* regime but has to be always attacked in *fortissimo.*

V. CONCLUSION

A differential formulation has been presented for a general bore geometry which allows the systematic study of the system stability at the threshold of oscillation. It gives a simple access to the effective dynamical description of a woodwind through a linear approach at the neighborhood of the equilibrium position. The parameters needed in differential description are deduced from the real bore impedance curve.

The application of this formulation to a particular case has proved to be efficient. Several time-domain numerical simulations have also been carried out to explore the behavior far from the threshold of oscillation, and an unexpected result has been obtained: It is necessary to increase the blowing pressure to relatively high values in order to leave the starting quasi-periodic oscillation and attain a stable periodic one. While traditionally it has been accepted that the evolution of the system behavior as the blowing pressure increases goes from equilibrium to multiphonic sound through a periodic one, our results show an evolution inverse in some way to that one, as the periodic sound is attained from the equilibrium through the multiphonic sound. Even if this result was already known from musical practice, it is not included in the classical frame of musical acoustics of woodwinds.

For different formulations concerning the excitatory device, different results may surely be obtained. Our purpose has just been to explore the validity of the presented approach for a particularly simple formulation.

Agulló, J., and Barjau, A. (**1986**). ''The reflection function $r(t)$: a matrix approach versus FFT,'' J. Sound Vib. **106**(2), 193–201.

Barjau, A. (**1987**). ''Contribució a l'estudi de l'acústica dels instruments de canya de la cobla,'' Doctoral thesis, University of Barcelona.

Barjau, A., Cardona, S., and Jordi, L. (**1990**). ''Fast simulation of the internal pressure in a double-reed woodwind,'' Coll. Mod. Phys. (Grenoble) **13**, 161–178.

Barjau, A., Gibiat, V., and Grand, N. (**1997**). ''Study of woodwind-like systems through nonlinear differential equations. Part I. Simple geometry,'' J. Acoust. Soc. Am. **102**, 3023–3031.

Benade, A. H. (**1976**). *Fundamentals of Musical Acoustics* (Oxford U.P., New York), see Chap. 21.

Gibiat, V. (**1988**). ''Phase space representations of acoustical musical signals,'' J. Sound Vib. **123**, 529–536.

Grand, N. (**1994**). ''Etude du seuil d'oscillation des systèmes acoustiques non-linéaires de type instrument à vent,'' Doctoral thesis, University of Paris VII.

McIntyre, M., Schumacher, R., and Woodhouse, J. (**1983**). ''On the oscillations of musical instruments,'' J. Acoust. Soc. Am. **74**, 1325–1345.

Manneville, P. (**1990**). *Dissipative Structures and Weak Turbulence* (Academic, New York).

Worman, W. E. (**1971**). ''Self-sustained non-linear oscillations of medium amplitude in clarinetlike systems,'' Ph.D. dissertation, Dept. of Physics, Case Western Reserve University, Cleveland, OH.

Transmission line method for the measurement of the acoustic nonlinearity parameter in biological liquids at very high frequencies

J. Kushibiki, M. Ishibashi, N. Akashi, T. Sannomiya, N. Chubachi, and F. Dunn[a)]
Department of Electrical Engineering, Tohoku University, Sendai 980-77, Japan

(Received 18 January 1997; accepted for publication 5 August 1997)

Nonlinear wave propagation in a system consisting of a liquid specimen held between two SiO_2 buffer rods is studied as a new method of determining the acoustic nonlinearity parameter of liquids at very high frequencies. Since the sign of the nonlinearity parameter of SiO_2 is opposite that for liquids, the system can be viewed as a transmission line containing regions of both positive and negative nonlinearity. The nonlinear characterization curve is employed in which the second harmonic component output of the receiving transducer is plotted as a function of propagation distance in liquids between the two SiO_2 buffer rods. The nonlinearity parameter is determined by measuring the dip position at which the transducer output becomes zero in the curve. Measurements are reported for water and aqueous solutions of dextrose and dextran in the fundamental frequency range 100–200 MHz. © *1997 Acoustical Society of America.* [S0001-4966(97)05811-6]

PACS numbers: 43.80.Ev, 43.80.Cs, 43.25.Zx, 43.35.Bf

INTRODUCTION

For studies of ultrasonic tissue characterization, a bioultrasonic spectroscopy system and method operating in the VHF and UHF ranges has been developed.[1] Bulk linear acoustic properties of velocity, attenuation coefficient, impedance, and density can be determined simultaneously using an ultrasonic transmission line in which the specimen is sandwiched between two synthetic silica glass (SiO_2) buffer rods having ZnO film ultrasonic transducers on their outer ends, and with distilled water employed as the reference medium. Measurements of some liquids and biological media such as the bovine tissues of liver, heart muscle, and fat have been reported in the frequency range from 10 to 500 MHz, revealing the significantly different properties from those measured at the low megahertz diagnostic medical frequencies.[1,2] We have taken a strong interest in nonlinear propagation phenomena in liquids and liquidlike media such as tissues at these high frequencies, as the acoustic nonlinearity parameter B/A has been extensively investigated for tissue characterization at low megahertz medical frequencies,[3–11] and have desired to accumulate basic data of biological media with the nonlinear acoustic parameter as well as with the linear acoustic parameters in this frequency range, for a new research area of tissue characterization.

The broadband operation of ZnO film ultrasonic transducers fabricated on SiO_2 buffer rods, using radio frequency (rf) tone burst pulses, is taken as the great advantage for measurements of frequency characteristics of the linear acoustic parameters. In order to apply the transmission line methodology to measure the acoustic nonlinearity parameter of liquids and liquidlike media, such as tissues, using the finite amplitude method, some problems result from the nonlinear generation in the two SiO_2 buffer rods, for which the nonlinearity parameter has a negative value, in contradistinction to the nonlinearity parameters of liquids/tissues, and most of other solid materials, reported as positive values.[12–14] In this paper, nonlinear wave propagation in the SiO_2/liquid–specimen/SiO_2 system, in the fundamental frequency range 100–200 MHz, is investigated theoretically and experimentally for an input fundamental-frequency ultrasonic power of about 10 mW. A new method of measuring the nonlinearity parameter of liquids is proposed, taking notice that the sign of the nonlinearity parameter of SiO_2 is different from that of liquids. Measurements are reported for water and 25% weight-fractional aqueous solutions of dextrose and dextran.

I. NONLINEAR PROPAGATION

A. Transducer output

A schematic view of the ultrasonic wave propagation is shown in Fig. 1 for the SiO_2/liquid–specimen/SiO_2 ultrasonic transmission line, where a liquid specimen is inserted between two SiO_2 buffer rods with ultrasonic transducers attached on their outer ends. As all materials are inherently nonlinear, generation of harmonics not only in the liquid but also in the SiO_2 buffer rods must be considered. In our analysis, only the second harmonic component is taken into account and all higher harmonics are ignored. The transmitting transducer is excited by electrical rf pulses at the fundamental frequency f_1 and emits ultrasonic rf pulses into the SiO_2 buffer rod. The receiving transducer detects ultrasonic rf pulses having the fundamental (f_1) and the second harmonic ($f_2 = 2f_1$) components.

In order to describe nonlinear wave propagation in this system, we consider the particle velocities for the fundamental and second harmonic components. Assuming finite, but moderate amplitude wave propagation within the discontinu-

[a)]Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois, Urbana, IL 61801.

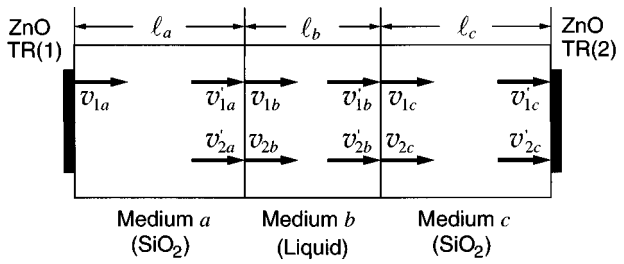


FIG. 1. A model of wave propagation in the SiO_2/liquid–specimen/SiO_2 ultrasonic transmission line, where v_1 and v_2 represents the particle velocities for the fundamental and second harmonic components.

ity distance, the second harmonic particle velocity $v_2(z)$ at the distance z from a fundamental sound source with the particle velocity $v_1(0)$ can be, in general, given by[3,15,16]

$$v_2(z)=\frac{\beta\pi f_1}{2c_0^2}\,v_1^2(0)\times\frac{\exp(-\alpha_2 z)-\exp(-2\alpha_1 z)}{2\alpha_1-\alpha_2}\,\mathrm{DIFF}(z), \tag{1}$$

where β is the nonlinearity parameter in the medium, c_0 is the infinitesimal wave velocity, α_1 and α_2 are the attenuation coefficients at f_1 and f_2, respectively, and DIFF(z) is the diffraction factor. The nonlinearity parameters for liquids, β_L, and isotropic solids, β_S, are defined by[14]

$$\beta_L=2+\frac{B}{A}, \tag{2}$$

and

$$\beta_S=-\frac{3C_{11}+C_{111}}{C_{11}}, \tag{3}$$

where B/A is the ratio of the coefficient of the second degree term to that of the first degree term of the Taylor's series expansion representation of the equation of state relating the pressure and density,[17] and C_{11} and C_{111} are the second- and third-order elastic constants, respectively. For simplicity, to derive the particle velocities detected by the receiving transducer, we neglect the diffraction effect as DIFF(z)=1. In Fig. 1, each medium $i(=a,b,c)$ containing ultrasonic propagation loss is featured with the complex acoustic impedance Z_i and the medium by length l_i. The transmission coefficient T_{ij} for the particle velocity from medium i to medium j, where i and $j=a,b,c$ is defined as

$$T_{ij}=\frac{2Z_i}{Z_i+Z_j}. \tag{4}$$

The particle velocity of the fundamental component, v'_{1c}, received by the transducer following transmission through the three media is simply obtained as

$$v'_{1c}=v_{1a}T_{ab}T_{bc}\,\exp[-(\alpha_{1a}l_a+\alpha_{1b}l_b+\alpha_{1c}l_c)], \tag{5}$$

where v_{1a} is the arbitrary amplitude of the particle velocity at the transmitting transducer. The particle velocity of the second harmonic component at the receiving transducer can be obtained as

$$\begin{aligned}v'_{2c}=&T_{ab}T_{bc}K_a v_{1a}^2\mathrm{ATT}_a\,\exp[-(\alpha_{2b}l_b+\alpha_{2c}l_c)]\\&+T_{bc}K_b[T_{ab}v_{1a}\,\exp(-\alpha_{1a}l_a)]^2\mathrm{ATT}_b\\&\times\exp(-\alpha_{2c}l_c)+K_c\{T_{ab}T_{bc}v_{1a}\\&\times\exp[-(\alpha_{1a}l_a+\alpha_{1b}l_b)]\}^2\mathrm{ATT}_c,\end{aligned} \tag{6}$$

where $K_i=\beta_i\pi f_1/c_{0i}^2$ and ATT_i is defined by

$$\mathrm{ATT}_i=\frac{\exp(-\alpha_{2i}l_i)-\exp(-2\alpha_{1i}l_i)}{2\alpha_{1i}-\alpha_{2i}}, \tag{7}$$

and $i=a,b,c$.

The fundamental frequency component, Eq. (5), is the same representation as that for the infinitesimal amplitude wave propagation. In Eq. (6), describing the particle velocity of the second harmonic v'_{2c}, the first term corresponds to the propagation of the second harmonic component generated in medium a, and the second term that in medium b, and the third term that in medium c, respectively.

B. Numerical calculation

Using Eq. (6), we made numerical calculations of the relative transducer outputs at the three fundamental frequencies 120, 150, and 180 MHz, for the second harmonic components in order to know the behavior of the SiO_2/water–specimen/SiO_2 system. The linear and nonlinear acoustic parameters of SiO_2 and water used in the calculations are shown in Table I. The length of the SiO_2 buffer rods is taken as 8 mm.

The results of the calculations are shown in Fig. 2. The relative transducer outputs vary with the length of the liquid medium l_b (propagation distance) between the two SiO_2 buffer rods and reach zero values at 577, 559, and 543 μm for the second harmonic frequencies of 240, 300, and 360 MHz, respectively. The dips result from the different signs of the nonlinearity parameters in SiO_2 and water (see Table I). It can be interpreted, therefore, that, before the dip position, the second harmonic components generated in the two SiO_2 buffer rods make greater contributions to the second harmonic component transducer output, whereas, beyond the dip position, those in the water with 180° phase difference in

TABLE I. Linear and nonlinear acoustic parameters used in the numerical calculations.

Sample	Velocity [m/s]	Attenuation α/f^2 [$\times10^{-15}$ s^2/m]	Density [$\times10^3$ kg/m^3]	β	Temperature [°C]
water	1496.7 (Ref. 18)	21.90 (Ref. 19)	0.9971 (Ref. 20)	7.11 (Ref. 21)	25
SiO_2	5954 (Ref. 22)	0.15 (Ref. 22)	2.202 (Ref. 22)	−11.58 (Refs. 12–14)[a]	23

[a]Temperature is 27 °C.

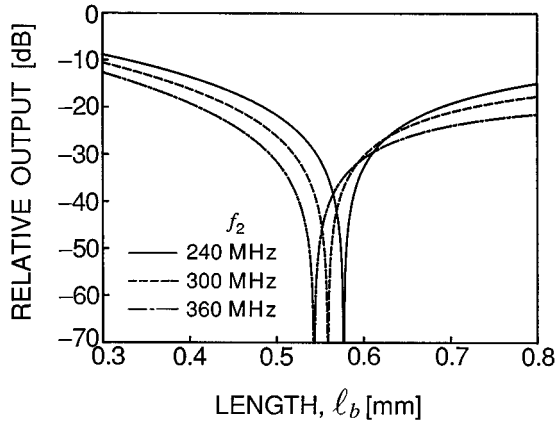


FIG. 2. Calculated receiving transducer outputs of the second harmonic components versus propagation distance in the SiO_2/water/SiO_2 system. The dip positions are 577, 559, and 543 μm for the second harmonic frequencies of 240, 300, and 360 MHz, respectively.

the nonlinearity parameter contribute more. And, as the frequency increases, the dip position appears at lesser values of liquid length. This is mainly due to the frequency dependence of attenuation of water. In this SiO_2/water/SiO_2 system, the ultrasonic waves are mostly attenuated by the water. Thus as the second harmonic component transducer output exhibits a characteristic curve featured with a dip at a particular propagation distance, the location of the dip in the nonlinear characterization curve can be employed for measurement of the nonlinearity parameter.

C. Determination of the nonlinearity parameter

As the media a and c are the same materials, SiO_2 with the negative nonlinearity parameter $\beta(SiO_2)$, and the medium b is the liquid specimen in which the propagation distance is alterable; in Eq. (6) the second harmonic component transducer outputs for the two SiO_2 buffer rods and liquid are canceled out at a certain propagation distance in the liquid $l_b = z_{d,}$, so that v'_{2c} becomes zero. Then, the nonlinearity parameter of the liquid specimen, β_L, is expressed as

$$\beta_L = -\frac{c_{0b}^2 \mathrm{ATT}_a \exp[-(\alpha_{2b} z_d + \alpha_{2c} l_c)]}{c_{0a}^2 T_{ab} \mathrm{ATT}_b \exp[-(2\alpha_{1a} l_a + \alpha_{2c} l_c)]}\beta(SiO_2) - \frac{c_{0b}^2 T_{ab} T_{bc} \mathrm{ATT}_c \exp[-2(\alpha_{1a} l_a + \alpha_{1b} z_d)]}{c_{0c}^2 T_{ab} \mathrm{ATT}_b \exp[-(2\alpha_{1a} l_a + \alpha_{2c} l_c)]}\beta(SiO_2). \tag{8}$$

Furthermore, when the length of two SiO_2 buffer rods are equal, i.e., $l_a = l_c$, the above equation can be rewritten as

$$\beta_L = -\frac{c_{0b}^2 \mathrm{ATT}_a\{\exp[-(\alpha_{2a} l_a + \alpha_{2b} z_d)] + T_{ab}(2 - T_{ab})\exp[-2(\alpha_{1a} l_a + \alpha_{1b} z_d)]\}}{c_{0a}^2 T_{ab} \mathrm{ATT}_b \exp[-(2\alpha_{1a} l_a + \alpha_{2a} l_a)]}\beta(SiO_2). \tag{9}$$

From these equations, we propose a new method of determining the nonlinearity parameter of the liquid in the space between the buffer rods by measuring the dip position z_d, with the premeasured/known linear acoustic parameters of the velocity, attenuation, and impedance, as well as with the known linear and nonlinear parameters of SiO_2 employed as the reference medium.

II. EXPERIMENTS

A. Measurement system

A block diagram of the measurement system is shown in Fig. 3. Measurements are carried out in rf pulse mode operation, and a signal generator and two bandpass filters, necessary for the nonlinear experiments, are introduced additionally to the pulse mode measurement system used in the bioultrasonic spectroscopy system.[1] This system consists of six main parts: the signal generator of electrical rf tone burst pulses, a pair of ultrasonic plane wave devices including the liquid specimen, the mechanical system for alignment of the two ultrasonic devices and for translation to record the nonlinear characterization curve, the rf receiver, the temperature measurement system, and the computer to control the entire system and to process the measured data.

An rf pulse is passed through the bandpass filter tuned to the fundamental frequency used in experiments, in order to

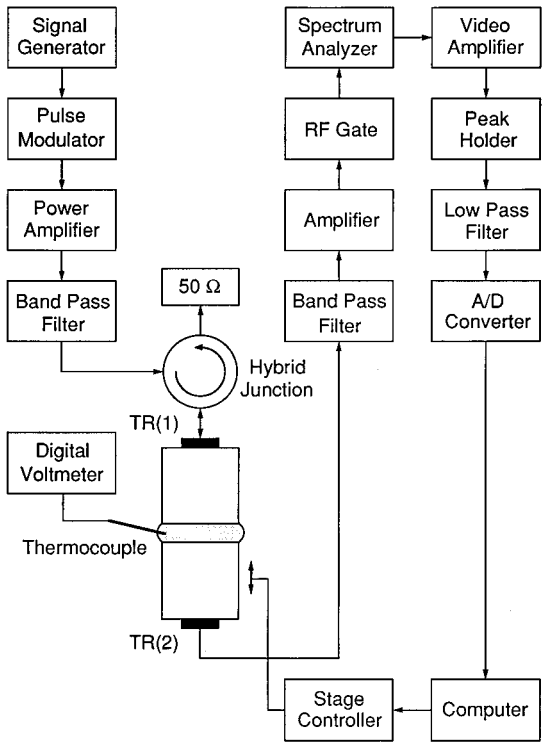


FIG. 3. Block diagram of the measurement system.

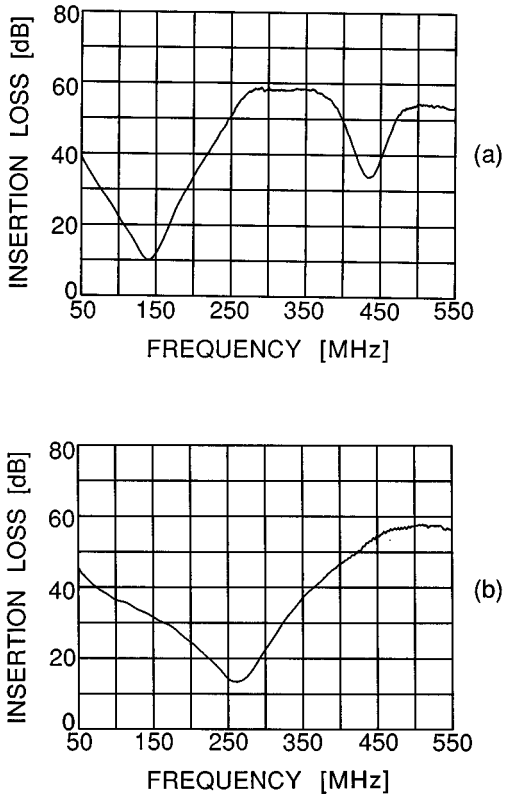


FIG. 4. Frequency characteristics of the insertion loss of the transmitting and receiving ultrasonic devices. (a) Transmitter; (b) receiver.

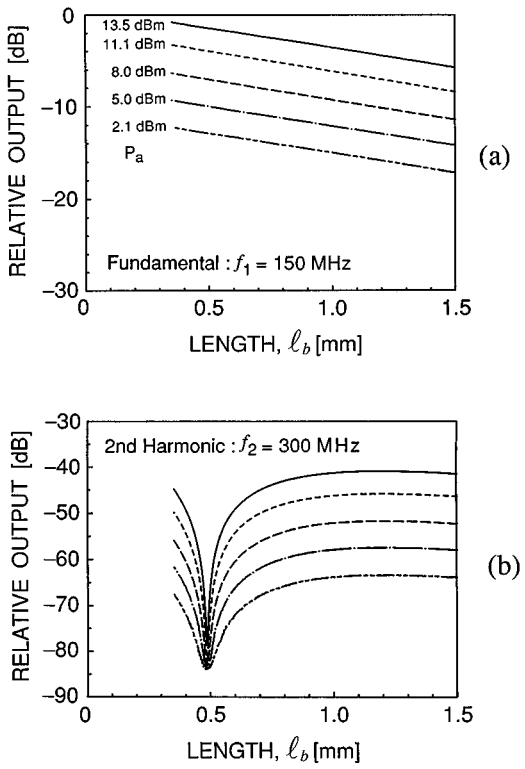


FIG. 5. Receiving transducer outputs for water as the liquid between the buffer rods versus propagation distance as a function of the acoustic power at the transmitting transducer. The dip position is 482 μm.

eliminate sufficiently its harmonics. The ultrasonic rf pulse wave, excited by the transmitting transducer TR(1), propagates through the two SiO_2 buffer rods and the liquid specimen, and the harmonics are generated in both media. The receiving transducer TR(2) detects a train of distorted ultrasonic rf pulse waves and converts them to electrical signals. The second bandpass filter selects the frequency components to be observed. The signals are amplified and the desired rf pulse signals are extracted by the rf gate circuit, and then detected with the spectrum analyzer. The A/D converted signals for the fundamental and second harmonic components are recorded as a function of the distance z by changing the spacing length between the two SiO_2 buffer rods by a z-axis translation stage with a stepping motor with a resolution of 0.1 μm. The temperature of the liquid specimen is measured by an inserted copper–constantan thermocouple.

The transducers used are of ZnO piezoelectric films fabricated by dc sputtering on the cylindrical SiO_2 (T-4040, Toshiba Ceramics Co.) rods 8 mm long and with diameters of 8 mm for the transmitter and 20 mm for the receiver, in order to support the liquid specimen in the gap between the two rods. The transmitting and receiving transducers were designed to have the film thicknesses of 19 μm around the center frequency of 150 MHz and of 9.5 μm around 300 MHz, respectively, having the same transducer diameter of 2.6 mm. Figure 4 shows the frequency characteristics of the insertion loss of the two devices. This transmitter can be used to excite ultrasonic waves at fundamental frequencies from 100 to 200 MHz, and the receiver can detect the corresponding fundamental and second harmonic components.

B. Results and discussions

1. Water

To confirm the nonlinear wave propagation measurement method described above, experiments were first conducted at 150 MHz for water for which the linear and nonlinear acoustic properties are well known.[18–21] Figure 5 shows the results of the receiving transducer output obtained as a function of acoustic power P_a at the transmitter. Measurements using the pulse width of 300 ns were carried out with the propagation distance larger than 0.3 mm. Figure 5(a) shows the results for the fundamental frequency and Fig. 5(b) for the second harmonic frequency. Both the transducer outputs at 150 and 300 MHz increase in proportion to the ultrasonic power used in experiments. The transducer output at 150 MHz decreases monotonically with the length of propagation distance in the liquid. The transducer output at 300 MHz becomes zero at the same propagation distance, as predicted in the calculated curves of Fig. 2. The dip distance was measured to be 482 μm.

Next, similar experiments were carried out for two other fundamental frequencies, 120 and 180 MHz, with the ultrasonic power of 8 dBm. Figure 6 shows the results, together with those at 150 MHz, reflecting the frequency dependence of attenuation. The dips were observed clearly at the different propagation distances, viz., at 496, 482, and 464 μm for the fundamental frequencies of 120, 150, and 180 MHz, respectively, as theoretically predicted above.

From the dip positions obtained above, the nonlinear parameters for water were calculated, using Eq. (9) with the

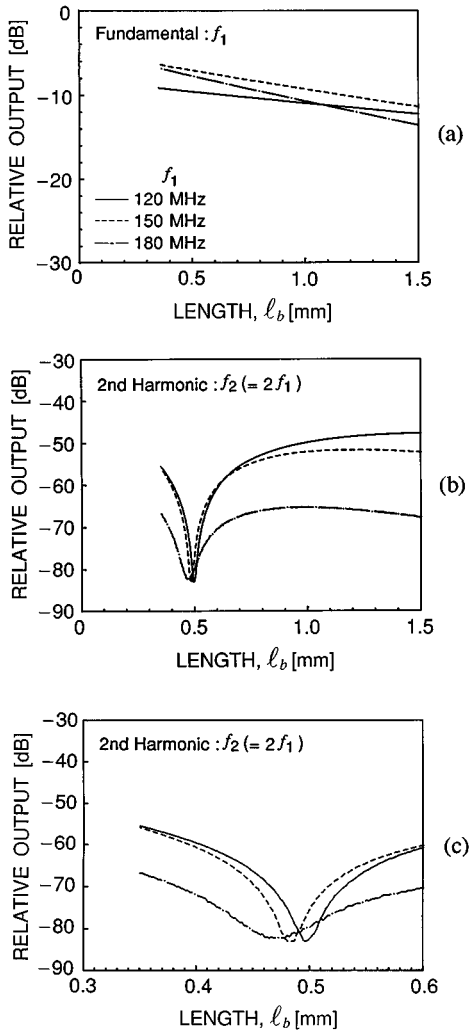


FIG. 6. Receiving transducer outputs for water versus propagation distance as a function of the fundamental frequency. The dip positions are 496, 482, and 464 μm for the second harmonic frequencies of 240, 300, and 360 MHz, respectively.

relevant acoustic parameters of water and SiO_2 given in Table I, and compared with published data measured at 9.9 MHz,[21] as shown in Table II. Slightly different β_W values of 8.34–8.44 were obtained, depending on the frequencies employed, but approximately 20% larger than the published data.

2. Aqueous solutions of dextrose and dextran

Twenty-five percent weight-fractional aqueous solutions of dextrose and dextran of different molecular weights were

TABLE II. Determined nonlinearity parameter for water.

f_1	β_W
120 MHz	8.36
150 MHz	8.34
180 MHz	8.44
9.9 MHz	7.11 (Ref. 21)[a]

[a]$B/A=5.11$ at 25 °C.

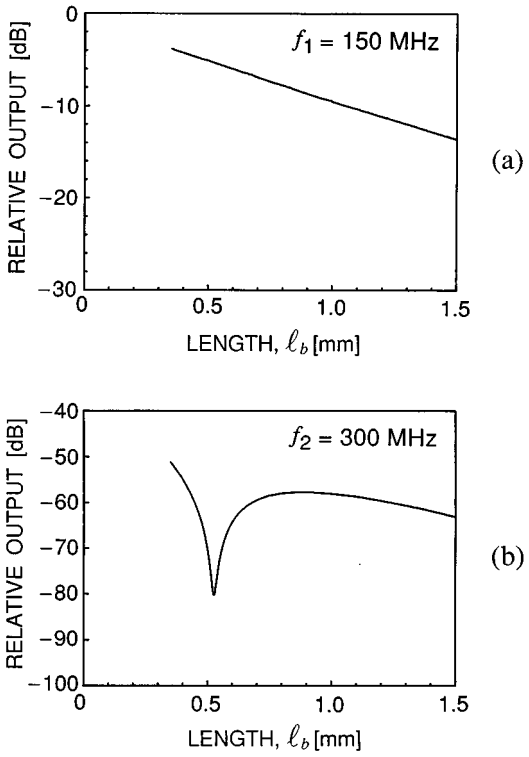


FIG. 7. Receiving transducer outputs for 25% aqueous dextrose solution versus propagation distance. The dip position is 526 μm.

studied at the fundamental frequency of 150 MHz. The molecular weights were 180 Daltons for the dextrose and $(2-3)\times10^5$ Daltons for the dextran solutes. The measured results of the transducer outputs are shown in Fig. 7 for 25% aqueous dextrose solution and in Fig. 8 for 25% aqueous dextran solution. It is seen from these data that the 25% dextrose aqueous solution has the linear acoustic properties of lesser attenuation, and the slightly larger dip position in the nonlinear characterization curve. To determine their nonlinear parameters, we also measured the linear acoustic parameters of velocity, attenuation, and impedance,[1] and calculated the nonlinearity parameters, comparing them with the published data,[5] as shown in Table III. The nonlinearity parameter of 9.18 for the 25% aqueous dextrose solution is smaller than the 9.56 value for the 25% aqueous dextran solution. Both values are approximately 20% larger than the published data.

The relative nonlinearity parameter, $\beta_R(=\beta_L/\beta_W)$, obtained by dividing the measured values by the nonlinearity parameter value of water, β_W, actually obtained in the same system at the same frequency, might lead to more reasonable results based on the two following reasons: (1) water is one of the better known and most physically and chemically stable media and one of the most important constituents of biological media; and (2) measurement errors due to the total diffraction effects in this system might be experimentally minimized because the velocities of the liquids are close to the velocity of water, as previously introduced in the bioultrasonic spectroscopy system.[1] For the 25% aqueous solutions of dextrose and dextran, we obtain 1.10 and 1.15, respectively, as given in Table III.

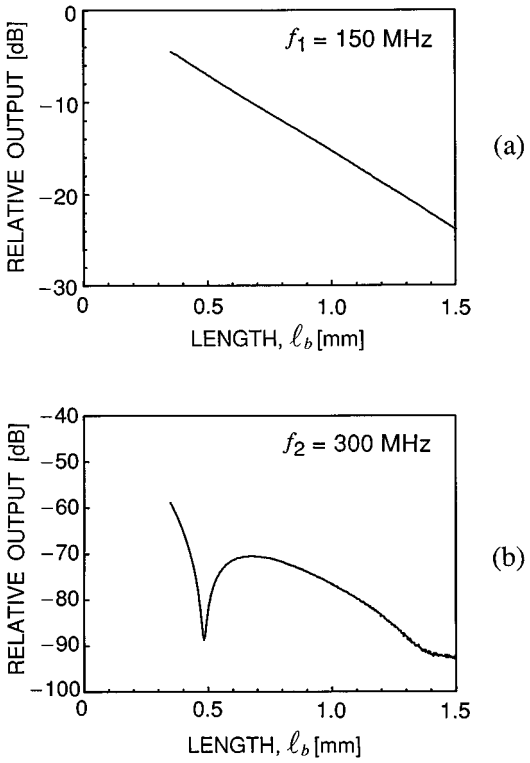


FIG. 8. Receiving transducer outputs for 25% aqueous dextran solution versus propagation distance. The dip position is 487 μm.

From these experiments, we could find out the first new information that the nonlinearity parameter is dependent on the molecular weight when these very high frequencies are utilized, although it was reported that the nonlinearity parameter in the dextrose and dextran aqueous solutions is insensitive to molecular weight and nearly the same for the same concentration, obtained by the finite amplitude and thermodynamic methods.[5]

C. Measurement accuracy

Accuracy of the nonlinearity parameters measured for the 25% aqueous solutions of dextrose and dextran with this prototype system, carried out in an air-conditioned room but wherein the temperature of specimen is measured independently, is discussed using Eq. (9), for which the diffraction effect is ignored. The measurement accuracy depends on the measurement of the dip position z_d in the nonlinear characterization curve, the measurement of the linear acoustic parameters of velocity, attenuation, and impedance for the aqueous solutions, the values of the linear and nonlinear acoustic parameters for the SiO_2 buffer rods as the reference medium, and the measurement of the length of the buffer rods. We now assume sufficient accuracy for the necessary values for the SiO_2 rods. The length of buffer rods can be easily measured with better than $\pm 0.002\%$ by a length gauging system having a resolution of ± 0.1 μm. The reproducibility of about ± 0.5 μm in dip measurement corresponds to the error of $\pm 0.1\%$ in β. On the other hand, the measurement accuracies of the linear acoustic parameters for the aqueous solutions are estimated to be $\pm 0.05\%$ for velocity, $\pm 1\%$ for attenuation, and $\pm 1\%$ for impedance in this study, which contribute to the errors in β of $\pm 0.1\%$, $\pm 0.02\%$, and $\pm 0.6\%$, respectively. It can be understood that the measurement error of the acoustic impedance has the greatest influence on the β determination. From these considerations, the measurement accuracy in the present measurements, neglecting diffraction effects, is estimated to be around $\pm 0.8\%$.

The significant differences between the previously published values of β and those measured in this study for water and the aqueous solutions might be attributed to neglect of the diffraction effects in the mathematical treatment. This suggests that further theoretical and experimental studies on the diffraction correction are very important. As the relative measurement error is now estimated to be $\pm 0.8\%$ in β, the difference of 4% obtained between the two 25% aqueous solutions of dextrose and dextran is meaningful and it is considered to result from the different molecular weights.

III. CONCLUDING REMARKS

Nonlinear wave propagation in the system of SiO_2/liquid–specimen/SiO_2 has been investigated theoretically and experimentally in the frequency range of the fundamental from 100 to 200 MHz, taking notice that the sign of the nonlinearity parameter of SiO_2 is different from that of liquids. It has been observed that, in the nonlinear characterization curve which records the second harmonic component output of the receiving transducer as a function of propagation distance in the liquid specimen between two SiO_2 buffer rods, there exists a dip at a certain distance z_d where the receiver transducer output becomes zero. A new method of determining the nonlinearity parameter of liquids, by measuring the dip position and the acoustic parameters of velocity, attenuation, and impedance with reference to SiO_2 with well-known linear and nonlinear acoustic parameters, has been developed; measurements have been made with water and aqueous solutions of dextrose and dextran.

Further studies on the effect of diffraction on the determination of the nonlinearity parameters will complete the method, and the following problems will have to be investigated at these very high frequencies: effect of velocity dis-

TABLE III. Determined linear and nonlinear acoustic parameters for aqueous solutions of dextrose and dextran. $\beta_W = 8.34$ at fundamental frequency of 150 MHz.

Sample	Velocity [m/s]	Attenuation $\alpha[\times 10^3$ neper/m] (150 MHz)	Attenuation $\alpha[\times 10^3$ neper/m] (300 MHz)	Density [$\times 10^3$ kg/m^3]	β_L	β_L/β_W	β_L (Publ.)
25% Dextrose	1623	0.847	3.31	1.06	9.18	1.10	7.96 (Ref. 5)
25% Dextran	1587	1.34	4.18	1.12	9.56	1.15	8.05 (Ref. 5)[a]

[a]Concentration is 24%.

persion on measurements, measurement limitation for media with very high ultrasonic losses, and modification of the method suitable for semi-solid biological tissues having a predetermined thickness. This method will be able to yield more accurate measurements than the conventional finite amplitude and thermodynamic methods because absolute wave amplitude measurements are replaced by length measurements.

Bioultrasonic research in the VHF and UHF ranges will be initiated in the near future with the bioultrasonic spectroscopy system with both functions of measuring the linear and nonlinear acoustic properties of biological media.

ACKNOWLEDGMENT

This research was supported in part by a Research Grant-in-Aid of the Ministry of Education, Science, and Culture of Japan.

[1] J. Kushibiki, N. Akashi, T. Sannomiya, N. Chubachi, and F. Dunn, ''VHF/UHF range bioultrasonic spectroscopy system and method,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 1028–1039 (1995).

[2] N. Akashi, J. Kushibiki, N. Chubachi, and F. Dunn, ''Acoustic properties of selected bovine tissues in the frequency range 20–200 MHz,'' J. Acoust. Soc. Am. **98**, 3035–3039 (1995).

[3] F. Dunn, W. K. Law, and L. A. Frizzell, ''Nonlinear ultrasonic wave propagation in biological materials,'' Proceedings of the IEEE Ultrasonics Symposium 527–532 (1981).

[4] W. K. Law, L. A. Frizzell, and F. Dunn, ''Ultrasonic determination of the nonlinearity parameter B/A for biological media,'' J. Acoust. Soc. Am. **69**, 1210–1212 (1981).

[5] W. K. Law, L. A. Frizzell, and F. Dunn, ''Determination of the nonlinearity parameter B/A of biological media,'' Ultrasound Med. Biol. **11**, 307–318 (1985).

[6] C. M. Sehgal, G. M. Brown, R. C. Bahn, and J. F. Greenleaf, ''Measurement and use of acoustic nonlinearity and sound speed to estimate composition of excised livers,'' Ultrasound Med. Biol. **12**, 865–874 (1986).

[7] J. Zhang, M. S. Kuhlenschmidt, and F. Dunn, ''Influences of structural factors of biological media on the acoustic nonlinearity parameter B/A,'' J. Acoust. Soc. Am. **89**, 80–91 (1991).

[8] C. M. Sehgal, R. C. Bahn, and J. F. Greenleaf, ''Measurement of the acoustic nonlinearity parameter B/A in human tissues by a thermodynamic method,'' J. Acoust. Soc. Am. **76**, 1023–1029 (1984).

[9] R. L. Errabolu, C. M. Sehgal, and J. F. Greenleaf, ''Dependence of ultrasonic nonlinearity parameter B/A on fat,'' Ultrason. Imaging **9**, 180–194 (1987).

[10] J. Zhang, ''Influences of biological structural features on the acoustic nonlinearity parameter B/A,'' Ph.D. dissertation, University of Illinois at Urbana–Champaign (1990).

[11] E. C. Everbach and R. E. Apfel, ''An interferometric technique for B/A measurement,'' J. Acoust. Soc. Am. **98**, 3428–3438 (1995).

[12] J. H. Cantrell, Jr. and M. A. Breazeale, ''Ultrasonic investigation of the nonlinearity of fused silica for different hydroxyl-ion contents and homogeneities between 300 and 3 K,'' Phys. Rev. B **17**, 4864–4870 (1978).

[13] M. A. Breazeale and J. Philip, ''Determination of third-order elastic constants from ultrasonic harmonic generation measurements,'' in *Physical Acoustics*, Vol. XVII, edited by W. P. Mason and R. N. Thurston (McGraw-Hill, New York, 1984), pp. 1–60.

[14] M. A. Breazeale, ''Physics and Engineering Principles of Nonlinear Acoustics,'' Jpn. J. Appl. Phys. **27**, Suppl. 27–1, 12–16 (1988).

[15] A. L. Thuras, R. T. Jenkins, and H. T. O'Neil, ''Extraneous frequencies generated in air carrying intense sound waves,'' J. Acoust. Soc. Am. **6**, 173–180 (1935).

[16] F. Ingenito and A. O. Williams, Jr., ''Calculation of second-harmonic generation in a piston beam,'' J. Acoust. Soc. Am. **49**, 319–328 (1971).

[17] R. T. Beyer, ''Nonlinear propagation in fluids,'' in *Nonlinear Acoustics* (Naval Ship Systems Command, New London, CT, 1974), Chap. 3, pp. 91–164.

[18] W. Kroebel and K.-H. Mahrt, ''Recent results of absolute sound velocity measurements in pure water and sea water at atmospheric pressure,'' Acustica **35**, 154–164 (1976).

[19] J. M. M. Pinkerton, ''The absorption of ultrasonic waves in liquids and its relation to molecular constitution,'' Proc. Phys. Soc. London, Sect. B **20**, 129–141 (1949).

[20] G. S. Kell, ''Density, thermal expansivity, and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale,'' J. Chem. Eng. Data **20**, 97–105 (1975).

[21] Z. Zhu, M. S. Roos, W. N. Cobb, and K. Jensen, ''Determination of acoustic nonlinearity parameter B/A from phase measurement,'' J. Acoust. Soc. Am. **74**, 1518–1521 (1983).

[22] Our data measured for SiO_2 (T-4040, Toshiba Ceramics Co., Japan).

LETTERS TO THE EDITOR

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

Systems approach to sound in porous media

Robert F. Lambert
Department of Electrical Engineering, Institute of Technology, University of Minnesota, Minneapolis, Minnesota 55455

(Received 27 February 1997; accepted for publication 14 July 1997)

A method based on linear systems analysis is developed for estimating propagation characteristics of sound in porous media both in the fluid of the pores and in the elastic structural frame. It predicts propagation constants, pressure ratios, and velocity ratios in close agreement with previous work. The two coupled systems are visualized in the form of signal flow diagrams that are composed of transfer admittances and transfer impedances. © *1997 Acoustical Society of America.* [S0001-4966(97)00811-4]

PACS numbers: 43.20.Bi, 43.20.Hq, 43.20.Tb, 43.35.Mv [JEG]

INTRODUCTION

This Letter presents and discusses a linear systems approach to analyzing sound in highly porous acoustic materials saturated by a light fluid such as air. It considers dilatational coupled wave motion in both fluid and frame. The objective is to formulate a simplified set of coupled equations of motion that facilitates analytical estimates of the desired propagation constants.

The basic fourth-order coupled system as formulated in previous work[1] is decomposed into two second-order coupled systems that are easily visualized and analyzed. The fourth-order system is shown schematically in Fig. 1 and the two second-order systems are shown in Fig. 2. All of the systems parameters are defined and listed in the Appendices.

A somewhat similar scheme was used in Refs. 2 and 3, where the frame was a fine fiber, layered material that had bulk elastic properties allowing internal fiber resonances. Comparisons between theoretical predictions of the effective internal impedance of the fluid pores and experimental measurements were very good over a wide range of frequencies, especially very low frequencies below resonance. The experimental techniques employed were discussed in Ref. 4. Previous work in Refs. 1–3 also verified that this model gave high-frequency estimates in good agreement with the rigid frame model. However, this is not the case at low frequencies (or in the vicinity of internal fiber resonance) where the rigid frame predictions of effective internal impedance can differ from measured values by up to an order of magnitude.

The notation used is identical with that employed in Refs. 1–3. Several new definitions of admittance and impedance transfer functions as depicted in Figs. 1 and 2 are also listed in the Appendices. Standard systems engineering notation has been used in the diagrams.

I. THEORETICAL ANALYSIS

The fourth-order linear coupled system is shown schematically in Fig. 1. The acoustic pressure amplitudes P_p and P_s and the velocity amplitudes V_p and V_s are complex numbers and the one-dimensional wave motion is proportional to $\exp[i(\omega t-\Gamma x)]$. The system in Fig. 1 is given in operational notation by a matrix equation of the form

$$\begin{vmatrix} (1-\Omega)Y_{11}\Gamma & Y_{12}\Gamma & -1 & 0 \\ (1-\Omega)Y_{21}\Gamma & Y_{22}\Gamma & 0 & -1 \\ -1 & 0 & z_{11}\Gamma & z_{12}\Gamma \\ 0 & -1 & z_{21}\Gamma & z_{22}\Gamma \end{vmatrix} \cdot \begin{vmatrix} P_s \\ P_p \\ V_p \\ V_s \end{vmatrix} = 0, \quad (1)$$

which can be derived from Eq. (1) in Ref. 1.

The system is formulated on a per unit length basis where parameters like $Y\Gamma$ and $(1-\Omega)\,Y\Gamma$ are acoustic admittance transfer functions and the $z\Gamma$ play the role of acoustic impedance transfer functions. They depend on the physical properties of both fluid and frame as indicated in the appendicies. The characteristics of Γ for a fine fiber, layered, porous material were discussed in Ref. 1.

It is here proposed to decompose Eq. (1) into two simpler second-order systems as given by Eqs. (2) and (3). This facilitates evaluating propagation constants of sound in both the fluid and frame in a fairly direct and reasonably accurate way. Thus two new propagation constants, Γ_p associated with the fluid pores and Γ_s associated with the structural frame, are hypothesized and incorporated into the original equations of motion such that $\Gamma_p P_p=\Gamma_s P_s$ over all space

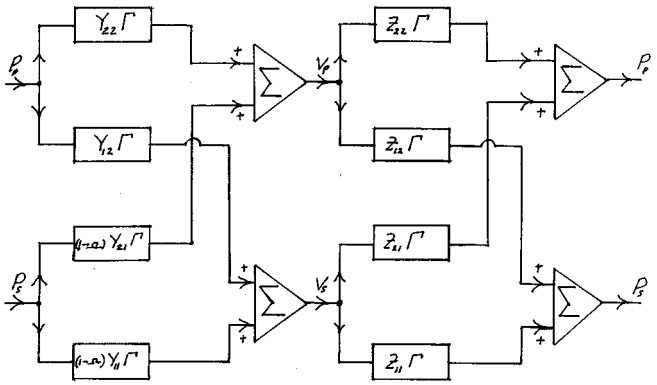


FIG. 1. Signal diagram representing the fourth-order coupled linear system expressed in Eq. (1). All symbols and element values are given in the appendicies.

and all frequencies. It is not required that $\Gamma_s=\Gamma_p$ and $P_s=P_p$ under all conditions, although they can be equal in limiting cases. In general $\Gamma_s\neq\Gamma_p$ and $P_s\neq P_p$. This hypothesis is accurate for many rather massive frames saturated with light fluids such as air.

The fourth-order system in Fig. 1 now can be decomposed into the two second-order systems as shown in Fig. 2. The system diagrams can be constructed in a number of ways and those shown require four transfer functions and one summer each. They then can be realized simply as two parallel paths.

The system in Fig. 2(a), which now represents sound in the fluid pores, is expressed as

$$\begin{vmatrix} Y_{T1}\Gamma_p & -1 & 0 \\ Y_{T2}\Gamma_p & 0 & -1 \\ -1 & z_{21}\Gamma_p & z_{22}\Gamma_p \end{vmatrix} \cdot \begin{vmatrix} P_p \\ V_s \\ V_p \end{vmatrix} = 0. \tag{2}$$

Only a single equation of state for the fluid and the appropriate Γ_p are used. The z_{21} and z_{22} parameters depend on the porosity of the frame and physical properties of the fluid only.

The system in Fig. 2(b), which now represents sound in the structural frame, is expressed as

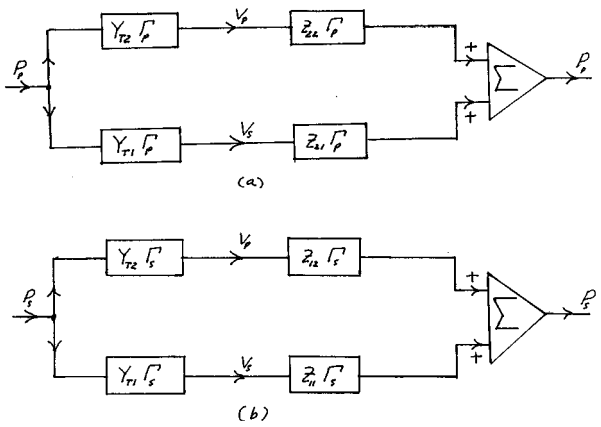


FIG. 2. Signal diagram representing the second-order coupled linear system expressed in Eqs. (2) and (3). (a) Sound in the fluid pores, Eq. (2). (b) Sound in the porous elastic frame, Eq. (3).

$$\begin{vmatrix} Y_{T1}\Gamma_s & -1 & 0 \\ Y_{T2}\Gamma_s & 0 & -1 \\ -1 & z_{11}\Gamma_s & z_{12}\Gamma_s \end{vmatrix} \cdot \begin{vmatrix} P_s \\ V_s \\ V_p \end{vmatrix} = 0. \tag{3}$$

Only a single equation of state for the frame and the appropriate Γ_s are used. The z_{11} and z_{12} parameters depend on the frame porosity and, in addition, the elastic properties at both fluid and frame, specifically the ratio κ_s/κ_p.

The analytical expressions for Γ_p, Eq. (2), and Γ_s, Eq. (3), are formulated as

$$\Gamma_p=[z_{21}Y_{T1}+z_{22}Y_{T2}]^{-1/2} \tag{4}$$

for propagation in the fluid, and

$$\Gamma_s=[z_{11}Y_{T1}+z_{12}Y_{T2}]^{-1/2} \tag{5}$$

for propagation in the elastic structural porous frame.

II. DISCUSSION AND OBSERVATION

Various characteristics of $\Gamma_p=\beta_p-i\alpha_p$, Eq. (4), were discussed in Refs. 2 and 3 with special regard to long, flexible, layered, fine fibers composing the frame. Predictions of the effective complex density of the fluid $\overline{\rho_p}=\Gamma_p^2/\omega\sigma_p$ were in good agreement with measured values over a wide range of frequencies obtained using techniques discussed in Refs. 4 and 5.

Presumably an effective complex frame density $\overline{\rho_s}=\Gamma_s^2/\omega\sigma_s$ could also be formulated, but, to date, no suitable measurement technique has been devised. Moreover, the quantitative behavior of κ_s which enters strongly into Eq. (5) has not been well-established for most flexible porous materials of interest in acoustics. Hence, even a numerical evaluation of $\Gamma_s=\beta_p-i\alpha_s$ is somewhat speculative as noted in Ref. 1. So, only a few simple asymptotic calculations and comparisons will be made here.

If the frame is very rigid (either massive or extremely stiff) Eq. (5) is analyzed in the form

$$\Gamma_s^2=(\sigma_s\Delta s)[\kappa_{12}S_{11}+(1-\Omega)\kappa_{12}S_{21}+\kappa_{11}S_{12}+(1-\Omega)\kappa_{11}S_{22}]^{-1}. \tag{6}$$

For the massive frame where S_{11} dominates, Eq. (6) reduces to

$$\Gamma_s^2=\sigma_sS_{22}/\kappa_{12}=\sigma_pS_{22}=\Gamma_p^2. \tag{7}$$

Hence, $P_s/P_p=\Gamma_p/\Gamma_s=1$ and $P_s=P_p$. This is a well-known result for massive frames.

In the extremely stiff frame case where $\kappa_s/\kappa_p\ll 1$ Eq. (5) yields a high-frequency asymptote

$$\Gamma_s \xrightarrow[\omega\to\infty]{} (\omega/c_{s\infty})(1-i\phi/(\omega\rho_{es})^{1/2}, \tag{8}$$

while the low-frequency prediction is

$$\Gamma_s \xrightarrow[\omega\to 0]{} (1-i)2^{-1/2}(\omega/c_{s0})(\phi_o/\omega\rho_{es})^{1/2}. \tag{9}$$

Equations (8) and (9) are in full agreement with asymptotic predictions made in Ref. 1.

For the stiff frame at low frequencies

$$P_s/P_p=\Gamma_p/\Gamma_s \xrightarrow[\omega\to 0]{} [\Omega/(1-\Omega)](\kappa_{p0}/\kappa_{s0})\gg 1, \tag{10}$$

which predicts a high-pressure ratio.

The velocity ratio V_s/V_p can be obtained from either Eq. (2) or Eq. (3) and formulated as

$$V_s/V_p=[(1-\Omega)S_{22}+S_{12}][S_{11}+(1-\Omega)S_{21}]^{-1}, \tag{11}$$

which depends only on mass and resistivity parameters. This ratio is highly frequency dependent, ranging in value from a maximum of $1/\Omega\geqslant 1$ at low frequencies to a minimum of $\rho_o/\Omega\rho_s\ll 1$ at extremely high frequencies. (See Ref. 2 for a more detailed discussion.)

It is concluded that this method of analyzing sound in porous frames saturated by a light fluid such as air yields results consistent with previous work. General conclusions regarding the behavior of Γ_s, Eq. (5), are difficult to draw based on present knowledge. More work on the vibration of elastic porous frames is required in order to obtain quantitative information. But, it is hoped that this method will be useful in certain cases, perhaps can be extended to other cases, and eventually implemented in design situations.

APPENDIX A: GLOSSARY OF SYMBOLS

α	attenuation constant
β	phase constant
Γ	$=\beta-i\alpha$, propagation constant
κ_p	inverse bulk modulus of fluid
κ_{p0}	static inverse bulk modulus of fluid
κ_s	inverse bulk modulus of porous frame
κ_{s0}	static inverse bulk modulus of porous frame
ϕ	dynamic resistivity of the fluid
ϕ_o	static resistivity of the fluid
ρ_a	apparent density of the attached mass
ρ_{es}	apparent density of the structural frame
ρ_s	mean density of the bulk porous frame
ρ_o	mean density of the fluid in free space
ρ_p	apparent density of fluid in pores
Ω	volume porosity
ω	angular frequency of excitation
c_p	speed of sound in pores
c_s	speed of sound in bulk porous frame
k_s	dynamic structure factor
P_p	acoustic pressure amplitude acting on fluid (phasor)
P_s	acoustic pressure amplitude acting on structural porous frame (phasor)
V_p	average fluid velocity amplitude (phasor)
V_s	average frame velocity amplitude (phasor)

APPENDIX B: FORMULAS AND DEFINITIONS

$$c_p=(\rho_p\Omega\kappa_p)^{-1/2}$$

$$c_s=(\rho_{es}\Omega^2\kappa_s)^{-1/2}$$

$$S_{11}=\Omega^2(\omega\rho_{es}-i\phi)$$

$$S_{12}=\Omega(\omega\rho_a-i\phi)=S_{21}$$

$$S_{22}=\omega\rho_p-i\phi$$

$$\rho_{es}=\rho_s(1-\Omega)/\Omega^{2^+}\rho_o(k_s-1)/\Omega$$

$$\rho_p=\rho_o k_s/\Omega$$

$$\rho_a=\rho_o(k_s-1)/\Omega$$

$$\sigma_p=\omega\kappa_p\Omega$$

$$\sigma_s=\omega\kappa_s(1-\Omega)$$

$$\kappa_{11}=1+[(1-\Omega)/\Omega]^2(\Omega\kappa_s/\kappa_p)$$

$$\kappa_{21}=1-\Omega$$

$$\kappa_{12}=[(1-\Omega)/\Omega^2](\Omega\kappa_s/\kappa_p)$$

$$\kappa_{22}=1$$

$$\Delta s=S_{11}S_{22}-S_{12}S_{21}$$

$$Y_{11}=S_{22}/\Delta s$$

$$Y_{21}=S_{21}/\Delta s=Y_{12}$$

$$Y_{22}=S_{11}/\Delta s$$

$$z_{11}=\kappa_{11}/\sigma_s$$

$$z_{12}=\kappa_{12}/\sigma_s$$

$$z_{21}=\kappa_{21}/\sigma_p$$

$$z_{22}=\kappa_{22}/\sigma_p$$

$$Y_{T1}=(1-\Omega)Y_{11}+Y_{12}$$

$$Y_{T2}=(1-\Omega)Y_{21}+Y_{22}.$$

[1] R. F. Lambert, ''Sound in a layered fine fiber porous material with finite frame stiffness,'' J. Acoust. Soc. Am. **84**, 1894–1905 (1988).

[2] R. F. Lambert, ''Acoustic resonance in highly porous, flexible, layered fine fiber materials,'' J. Acoust. Soc. Am. **93**, 1227–1234 (1993).

[3] R. F. Lambert, ''Low-frequency acoustic behavior of highly porous, layered, flexible, fine fiber materials,'' J. Acoust. Soc. Am. **97**, 818–821 (1995).

[4] J. D. McIntosh, M. T. Zuroski, and R. F. Lambert, ''Standing wave apparatus for measuring fundamental properties of acoustic materials in air,'' J. Acoust. Soc. Am. **88**, 1929–1938 (1992).

[5] M. D. Dahl, E. J. Rice, and D. E. Groesbeck, ''Effects of fiber motion on the acoustic behavior of an anisotropic, flexible fibrous material,'' J. Acoust. Soc. Am. **87**, 54–66 (1990).

Does air absorption modify ground effect?

Rufin Makarewicz[a)]
Kyushu Institute of Design, Department of Acoustic Design, 4-9-1 Shiobaru, Minami-ku, Fukuoka, Japan

(Received 25 April 1997; revised 23 June 1997; accepted 30 June 1997)

Sound propagation in the atmosphere involves many wave phenomena. Some of them, e.g., scattering on turbulence, and ground effect, interact with each other. Using the equation of state for relaxation in a binary mixture of gases, such as oxygen and water vapor, it is shown that the interaction between the ground effect and air absorption can be neglected. © *1997 Acoustical Society of America.* [S0001-4966(97)03810-1]

PACS numbers: 43.28.Fp, 43.50.Vt, 43.20.Fn [LCS]

INTRODUCTION

The models of outdoor noise propagation consider geometrical spreading, ground effect, air absorption, refraction, and scattering by atmospheric turbulence.[1] A mutual interaction occurs between these wave phenomena. For example, scattering disturbs the interference of the direct and reflected waves and, thus, diminishes the outcome of the ground effect.[2] The present study examines the interaction between air absorption and ground effect.

Reference 3 defines the expression for sound pressure above the ground in the atmosphere that can be rewritten in the form

$$p=A_0\left\{\frac{e^{i\tilde{k}R_1}}{R_1}+Q\left[\frac{\omega}{c}\right]\frac{e^{i\tilde{k}R_2}}{R_2}\right\}, \tag{1}$$

where R_1 and R_2 are the path lengths of the direct and reflected waves (Fig. 1), and the reflection coefficient is

$$Q=Q_1\left[\frac{\omega}{c}\right]+iQ_2\left[\frac{\omega}{c}\right]. \tag{2}$$

Note that the reflection coefficient, Q, depends upon $k=\omega/c$, i.e., the real part of the complex wave number $\tilde{k}$,

$$\tilde{k}=\frac{\omega}{c}+i\alpha. \tag{3}$$

Does this suggest that the ground effect is independent of air absorption as described by the coefficient α?

To answer this question, we make use of the Helmholtz equation for a gas with vibration relaxation, which is the most important mechanism for sound absorption in the atmosphere in the audible frequency range.[4]

I. THEORY

For cylindrical symmetry and harmonic time dependence, $\exp(-i\omega\cdot t)$, the Helmholtz equation is

$$\Delta p(R,z)+\tilde{k}^2 p(R,z)=0, \tag{4}$$

where the complex wave number is a function of frequency,

$$\tilde{k}=\frac{\omega}{cF(i\omega)}. \tag{5}$$

It is well known that the sound speed, c, is constant within the audible frequency range. In a special case of the Tuesday–Boudart approximation,[5] the function F is

$$F(i\omega)=\sqrt{\frac{1-i\omega\tilde{\tau}_\nu}{1-i\omega\tilde{\tau}_p}}. \tag{6}$$

Making use of the explicit form of $F(i\omega)$, we arrive again at Eq. (3), which we now write in the form

$$\tilde{k}=\frac{\omega}{c}\cdot(1+i\varepsilon). \tag{7}$$

Here, the coupling coefficient, ε, is defined by

$$\varepsilon=c\,\frac{\alpha(\omega)}{\omega}. \tag{8}$$

So far, obtained results show that in the absorbing atmosphere, above a ground surface, sound pressure, p, is determined by the Helmholtz equation [Eq. (4)], with the complex wave number, $\tilde{k}$ [Eq. (7)]. Equation (1) yields the solution to this equation, where $Q[\omega/c]$ is replaced by the reflection coefficient depending upon $\tilde{k}$,

$$Q=Q\left[\frac{\omega}{c}(1+i\varepsilon)\right]. \tag{9}$$

Does ε significantly change sound pressure?

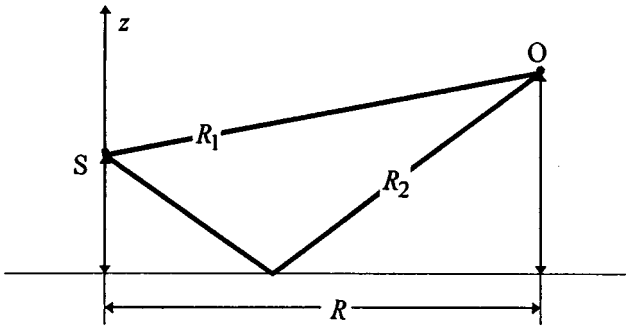


FIG. 1. Source, S, and receiver, O, geometry.

[a)]On leave from A. Mickiewicz University, Institute of Acoustics, Poznan, Poland.

II. CALCULATIONS

The atmospheric attenuation coefficient, $\alpha[Np/m]$, can be calculated from the standardized attenuation coefficient, $\tilde{\alpha}(\mathrm{dB/km})$, which embodies all processes of air absorption,[6]

$$\alpha = \frac{1}{2\lg(e)} \cdot 10^{-4}\tilde{\alpha}(f,T,h), \tag{10}$$

where $f(\mathrm{Hz})$ is frequency, $T(°\mathrm{C})$ denotes temperature, and $h(\%)$ expresses relative humidity.

For the temperature range, $-25<T<+50(°\mathrm{C})$, humidity range, $10<h<100(\%)$, and for two benchmark frequencies,[3] $f_1=100$ Hz and $f_2=1000$ Hz, the maximal values of the standardized attenuation coefficients are $\tilde{\alpha}_1=1.02$ (dB/km) and $\tilde{\alpha}_2=21.6$ (dB/km), respectively.[6] Equation (10) yields the corresponding attenuation coefficients, $\alpha_1\approx 1.17\times 10^{-4}$(Np/m) and $\alpha_2\approx 2.49\times 10^{-3}$ (Np/m). Finally, the maximal values of the coupling coefficient are as follows: $\varepsilon_1\approx 6.4\times 10^{-5}$ and $\varepsilon_2\approx 1.4\times 10^{-4}$ [Eq. (8)]. Within the audible frequency range the coupling coefficient is very small, $\varepsilon \ll 1$.

To quantify the impact of the change,

$$Q\left[\frac{\omega}{c}\right]\rightarrow Q\left[\frac{\omega}{c}(1+i\varepsilon)\right], \tag{11}$$

on sound propagation, we have calculated the sound-pressure level,

$$L_p = 10\log\{p^2/p_0^2\}, \quad p_0 = 20\ \mu\mathrm{Pa}, \tag{12}$$

where the mean squared sound pressure is

$$p^2 = A_0^2\left\{\frac{e^{-2\alpha R_1}}{R_1^2}+[Q_1^2+Q_2^2]\frac{e^{-2\alpha R_2}}{R_2^2} + 2\frac{e^{-\alpha(R_1+R_2)}}{R_1R_2}*\left[Q_1\cos\left(\frac{\omega}{c}(R_2-R_1)\right) + Q_2\sin\left(\frac{\omega}{c}(R_2-R_1)\right)\right]\right\}. \tag{13}$$

For $p^2(\varepsilon=0)$ with $Q[\omega/c]$ and $p^2(\varepsilon\neq 0)$ with the reflection coefficient, Q, depending on the coupling coefficient, ε [Eq. (8)], the difference of the sound-pressure levels is

$$\Delta L_p = 10\log\left\{\frac{p^2(\varepsilon=0)}{p^2(\varepsilon\neq 0)}\right\}. \tag{14}$$

Taking the benchmark frequencies, $f_1=100$ Hz, $f_2=1000$ Hz, and corresponding impedances, $Y_1=12.81$, $X_1=11.62$, and $Y_2=5.96$, $X_2=2.46$, we have calculated ΔL_p for the heights $h_s=5$ m and $h_0=1$ m, and for the horizontal distance, $R<10$ km (Fig. 1). In each case the difference of the sound-pressure levels is not significant: $\Delta L_p<0.01$ dB.

III. CONCLUSION

The Helmholtz equation [Eq. (4)], with the complex wave number, $\tilde{k}$ [Eq. (7)], describes both the phenomenon of sound absorption and reflection from the ground surface. A solution to Eq. (4) can be approximated by Eq. (1), where the reflection coefficient, $Q[\omega/c]$, is independent of the absorption coefficient, α. This lack of dependence indicates that the interaction between the ground effect and air absorption is negligible.

ACKNOWLEDGMENTS

The author is indebted to the anonymous reviewer and L. C. Sutherland (Associate Editor) for valuable comments.

[1] T. F. W. Embleton, ''Tutorial on sound propagation outdoors,'' J. Acoust. Soc. Am. **100**, 31–48 (1996).

[2] G. A. Daigle, ''Effects of atmospheric turbulence on the interference of sound waves above a finite impedance boundary,'' J. Acoust. Soc. Am. **65**, 45–49 (1979).

[3] K. Attenborough, S. Taherzadeh, H. E. Bass, X. Di, R. Raspert, G. R. Becker, A. Gudesen, A. Chrestman, G. A. Daigle, A. L'Esperance, Y. Gabillet, K. E. Gilbert, Y. L. Li, M. J. White, P. Naz, J. M. Noble, and H. A. J. M. van Hof, ''Benchmark cases for outdoor sound propagation models,'' J. Acoust. Soc. Am. **97**, 173–191 (1995).

[4] H. E. Bass, L. C. Sutherland, and A. J. Zuckerwar, ''Atmospheric absorption of sound: Update,'' J. Acoust. Soc. Am. **88**, 1019–1021 (1990); *ibid.* **97**, 680–683 (1995); *ibid.* **99**, 1259 (1996).

[5] L. W. Townsend and W. E. Meador, ''Vibration relaxation and sound absorption and dispersion in binary mixtures of gases,'' J. Acoust. Soc. Am. **99**, 920–925 (1996).

[6] ANSI S1.26-1995, ''Method for calculation of the absorption of sound by the atmosphere'' (American National Standards Institute, New York, 1995).

Further investigation on actively created quiet zones by multiple control sources in free space

Jingnan Guo and Jie Pan
Department of Mechanical and Materials Engineering, University of Western Australia, Nedlands, WA 6907, Australia

(Received 28 April 1997; accepted for publication 5 August 1997)

Further research on the quiet zones created in the free space, using multiple control sources and error microphones located in two parallel planes, indicated that there also exists a range of optimal spacing for the control sources and error microphones. Both the computer simulations and experimental results show that the control system can create the largest quiet zone and suffer the least increase of total sound power output when the control actuators and sensors are arranged within this optimal range. © *1997 Acoustical Society of America.* [S0001-4966(97)05411-8]

PACS numbers: 43.50.Ki [GAD]

INTRODUCTION

The authors have investigated the active control of local sound field in free space, which cancels the noise and creates the quiet zones in some desired places, and found that the performance of the control system is largely dependant on the configuration of the control system.[1] Their research on a typical control system with multiple control sources and error microphones equally placed in two parallel lines demonstrated that the spacings of the control sources and error microphones are vital to the performance of the control system. A range of optimal spacing of control sources and error microphones was discovered. When the control actuators and sensors are arranged within this range of spacing, the least increase in total power output and largest size of quiet zones will be resulted.[2]

If we extend the two parallel lines' arrangement of control sources and error microphones to the two parallel planes' arrangement, it is reasonable to believe that further increase of the size of quiet zones in 3-D space is possible.[3] The results presented in this paper show that there also exists an optimal range of spacing among the control sources and error microphones. The largest area of quiet zones and least increase of total power output can be achieved only when the condition of the optimal spacing is satisfied.

For the analyses presented here, a feedforward control mechanism has been assumed with a reference signal correlated to the output of the primary source. The control sources are modeled as constant volume velocity sources, the outputs of which are linearly related to their inputs.

I. SYSTEM DESCRIPTION

The arrangement of the control system is shown in Fig. 1. One point source was used as the primary noise source. Two grids of $N\times N$ monopole control sources and error microphones are placed in two parallel planes. The control source grid is r_{ps} away from the primary source, and r_{se} away from the error microphone grid. The separations between the two adjacent control sources and error microphones are r_{ss}. To minimize the primary sound pressures at the positions of $N\times N$ error microphones, the strengths of the $N\times N$ control sources in frequency domain should be chosen as

$$\mathbf{q}_s = -\mathbf{Z}_{se}^{-1}\mathbf{Z}_{pe}q_p, \tag{1}$$

where q_p is the strength of the primary source, $\mathbf{q}_s$ is a column vector of source strength for $N\times N$ control sources, $\mathbf{Z}_{se}$ is an $(N\times N)\times(N\times N)$ matrix of acoustic transfer impedance from the $N\times N$ control sources to the $N\times N$ error microphones, and $\mathbf{Z}_{pe}$ is a column vector of acoustic transfer impedance from the primary source to the $N\times N$ error microphones. Then the total radiated acoustic power of the system, which is the summation of power outputs of primary sources and secondary sources, can be written as

$$W_T = \tfrac{1}{2}\{|q_p|^2 Z_0 + \mathbf{q}_s^H\ \mathrm{Re}(\mathbf{Z}_{ss})\mathbf{q}_s + q_p^*\ \mathrm{Re}(\mathbf{Z}_{ps}^T)\mathbf{q}_s + q_p\mathbf{q}_s^H\ \mathrm{Re}(\mathbf{Z}_{ps})\}, \tag{2}$$

where $Z_0=\omega^2\rho_0/4\pi c_0$, $\mathbf{Z}_{ss}$ is an $(N\times N)\times(N\times N)$ transfer impedance matrix among the $N\times N$ control courses, and $\mathbf{Z}_{ps}$ is the column vector of transfer impedance between the primary source and the $N\times N$ control sources. The principle of acoustic reciprocity applies in this discussion, i.e., $\mathbf{Z}_{sp}=\mathbf{Z}_{ps}^T$.

II. NUMERICAL ANALYSIS

A. Total sound power output increase

For the local control system, the total sound power output always increases after control (when the primary source and control sources are at least one-half wavelength apart, this is always the case for local control). The increase of total power output of the control system is given by

$$\Delta W_T = 10\log(W_T/W_0)(\mathrm{dB}), \tag{3}$$

where W_T is the total sound power output of the control system expressed by Eq. (2), and W_0 is the sound power output of the primary source before the control.

The calculation of the total power output increase of the control system in different configurations showed that the

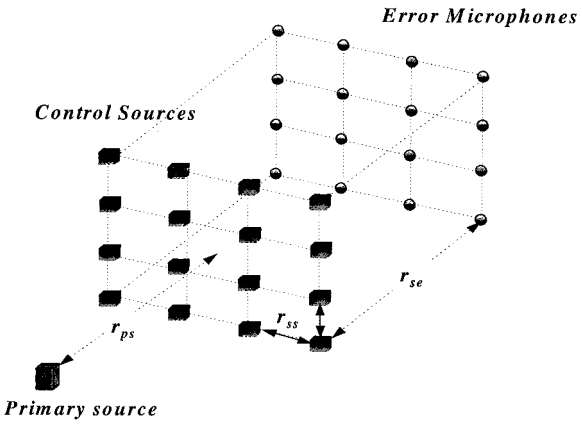


FIG. 1. The control system with multiple control sources when both control sources and error microphones are evenly placed in two parallel planes.

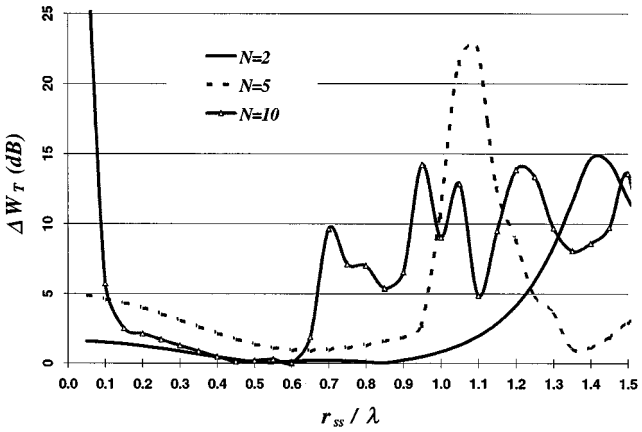


FIG. 2. The total power output increase of the control system for three configurations.

separation of the control sources always has a range with low increase of total power output. Figure 2 gives the examples of increase of total power output of the control system with the separation r_{ss} of the control sources (and error microphones), in which the distance from primary source to the control source plane $r_{ps}=\lambda$, and distance from control source plane to error microphone plane $r_{se}=2\lambda$, where λ is wavelength of the noise, and the numbers of the control sources (as well as the error microphones) are 2×2, 5×5, and 10×10 respectively.

It is obvious that there exists a range of separation distance between control sources r_{ss}, with which the total power output of the control system has the minimum increase showing in Fig. 2. The total power output of the control system increases greatly for the configuration outside this range. Using the relative level derivation $\Delta=r_{se}/(r_{se}+r_{ps})$, the upper and lower limits of the range associated with low power output increase (also called optimal range) can be accessed.[2] Figure 2 also shows that the upper limit decreases with the increasing number of the control sources. If the configurations are outside of the lower limit of the range, the increase of total power output can be extremely large when the number of control sources is large.

Figure 3 gives the examples of upper and lower limits of the optimal range for a control system of 5×5 control sources and error microphones with various arrangement of r_{ps} and r_{se}. It shows that both the upper and lower limits increase with the increasing distance between control sources grid and error microphone grid r_{se}. While the lower limit of the optimal range is proportional to the distance between primary source and control sources grid r_{ps}, the upper limit does not change with r_{ps}. Hence, the optimal range becomes narrower when r_{ps} is getting larger.

After the numerical calculation of the upper and lower limits of the optimal ranges for various configurations of the control systems with the control sources ranging from 2×2 to 20×20 are completed, the curve fitting of the upper and lower limits of the optimal range can be summarized as

$$r_{ss-\max}\cong\begin{cases}\dfrac{\lambda}{2}\sqrt{1+\dfrac{4r_{se}}{N\lambda}}, & N=2,4,6,...,\\ \dfrac{\lambda}{2}\sqrt{1+\dfrac{N+1}{N-1}\dfrac{4r_{se}}{N\lambda}}, & N=3,5,7,...,\end{cases}\tag{4}$$

and

$$r_{ss-\min}\cong\begin{cases}\dfrac{1.4\lambda}{20}\left[\ln\left(\dfrac{r_{ps}}{\lambda}+\dfrac{1}{2}\right)+0.5\right]\ln\left(\dfrac{r_{se}}{\lambda}+\dfrac{1}{2}\right)+\dfrac{\lambda}{3}, \\ \qquad N=2,4,6,...,\\ \dfrac{(N+1)\lambda}{20(N-2)}\left[1.25\ln\left(\dfrac{r_{ps}}{\lambda}+\dfrac{1}{2}\right)+0.7\right] \\ \qquad\times\ln\left(\dfrac{r_{se}}{\lambda}+\dfrac{N+1}{4(N-1)}\right)+\dfrac{\lambda}{3},\ N=3,5,7,... .\end{cases}\tag{5}$$

It can be seen that although the lower limit is fitted in a different form, the expression of the upper limit is the same as that of the control system with the control sources and error sensors placed in two parallel lines.[2] Similar to the case of control sources and error sensors in two parallel lines, it will be shown that the optimal configuration of the control system corresponding to a large quiet zone and small total power output increase is exactly within this range, i.e., $r_{ss-\min}\leqslant r_{ss}\leqslant r_{ss-\max}$.

B. Quiet zones

The ability to create a large area of quiet zone of the control system has been thoroughly investigated. It has been found that the control system can generate a large area of quiet zone when the configuration of the system is chosen within the optimal range. Although the control system can also create a large area of quiet zone with the separation of the control sources shorter than the lower limit of the optimal range, the sound pressure in the areas outside the quiet zone suffers a big increase. When the separation of the control sources is bigger than the upper limit of the optimal range, the control system is not able to create quiet zones, and the sound pressure also increases in most areas.

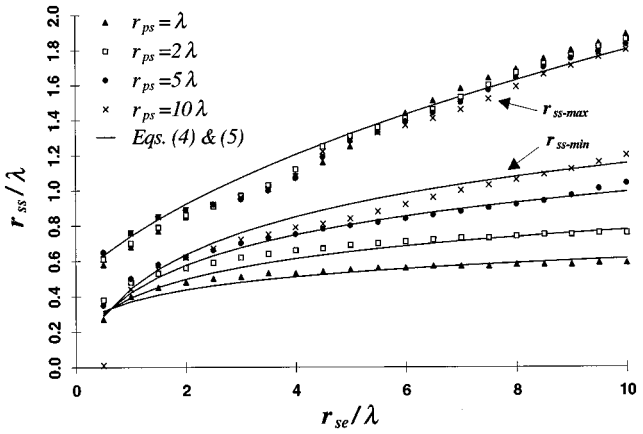


FIG. 3. The upper and lower limits of the control system with 5×5 control sources and error microphones.

Figure 4 is an example of quiet zones created by a control system with 5×5 control sources and error microphones, where the distance from primary source to the control source plane is $r_{ps}=\lambda$, and distance from control source plane to error microphone plane is $r_{se}=2\lambda$. The coordinates of the primary source, control sources, and error microphones are, respectively, (0,0,0), $((i-3)r_{ss},r_{ps},(j-3)r_{ss})$, and $((i-3)r_{ss},r_{se},(j-3)r_{ss})$ $(i=1,2,...,5;j=1,2,...,5)$. The upper limit and lower limit of the optimal range for this system can be calculated by Eqs. (4) and (5) as $r_{ss-\max}=0.92\lambda$ and $r_{ss}-\min=0.44\lambda$, respectively. Two separation distances of $r_{ss}=0.75\lambda$ as in Fig. 4(a) and $r_{ss}=1.08\lambda$ as in Fig. 4(b) are used for comparison.

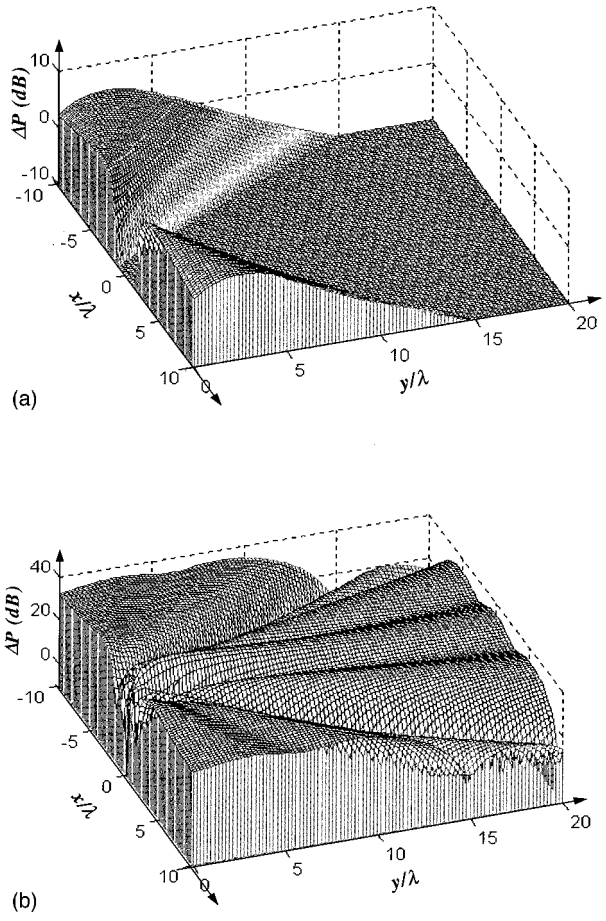


FIG. 4. Quiet zones created by a control system with 5×5 control sources and error microphones, when $r_{ss}=0.75\lambda$ as in (a) and $r_{ss}=1.08\lambda$ as in (b).

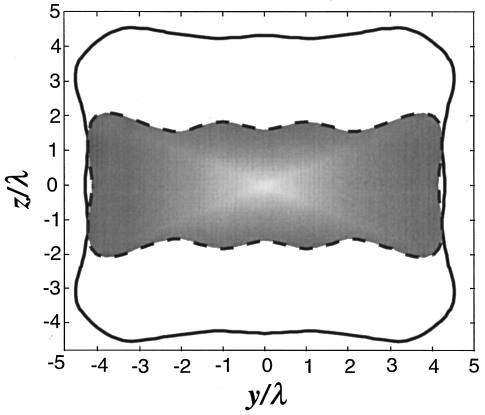


FIG. 5. Comparison of quiet zones created by the control system with multiple control sources evenly spaced in two parallel lines and that in two parallel planes.

Figure 4(a) shows that a large quiet zone is created by the system in the area behind the error microphones when the spacing of the control sources is within optimal range. When the spacing of the control sources is above the upper limit of the optimal range ($r_{ss}=1.08\lambda>r_{ss-\max}$), the system cannot create any obvious quiet zones, as shown in Fig. 4(b). A big increase of sound pressure in most places is also observed. This increase of sound pressure in some places can be as high as 30 dB!

Shown in Fig. 5 is an example of a comparison to demonstrate that the extension of control actuators and sensors from being evenly spaced in two parallel lines to being evenly spaced in two parallel planes further increases the size of the quiet zone. For the control system in Fig. 5, $N=9$, $r_{ps}=5\lambda$, and $r_{se}=10\lambda$. The optimal range for this configuration can be calculated as $0.71\lambda\leqslant r_{ss}\leqslant 1.28\lambda$. $r_{ss}=1.2\lambda$ is chosen for these two control systems, and the observation is conducted in a $y-z$ plane and 5λ away the error microphones. The area enclosed by the dashed line is the quiet zone created by the system with control sources and error microphones spaced in two lines, and the area enclosed by the solid line is the quiet zone created by the system with control actuators and sensors placed in two parallel planes.

It can be shown that the cross section of the quiet zone in $y-z$ plane is a rectangle for the system with control sources in a line, and a square with a width the same as that of the rectangle for the system with control sources in a plane. It is clear that the height of the quiet zone is significantly increased for the latter case.

III. EXPERIMENTAL VERIFICATIONS

Experiments were carried out in an anechoic chamber, 4.2 m× 4.2 m× 4.2 m in size. The simplest case, the control system with 2×2 control sources and the same number of error microphones, was tested. The primary noise source was a half-enclosed loudspeaker. Four half-enclosed speakers and four microphones were used as control sources and error

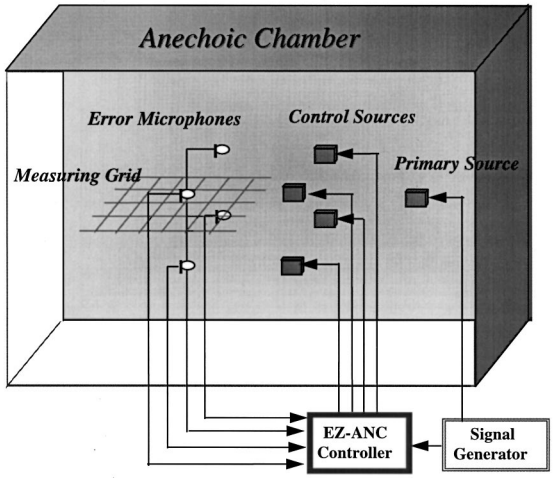


FIG. 6. Experimental setup in an anechoic chamber.

sensors respectively. A multi-channel EZ-ANC was applied as the controller. The sound signal used in the experiment was a pure tone of 500 Hz. This pure-tone signal was fed into the primary source directly, and was also provided to the controller as a reference signal. Four control channels of the controller were used to cancel the total sound pressure at the positions of four error microphones.

The arrangement of the control system is shown in Fig. 6, where both the four control sources and error microphones were placed in each corner of two parallel squares. The distance from the primary source to the control source plane is $r_{ps}=\lambda$, and the distance from the control source plane to error microphone plane is $r_{se}=2\lambda$. The coordinate of the primary source is set as $[0,-(r_{ps}+r_{se}),0]$, and the coordinates of the control sources and error microphones are $[(2i-3)r_{ss}/2,\ -r_{se},\ (2j-3)r_{ss}/2]$ and $[(2i-3)r_{ss}/2,\ 0,\ (2j-3)r_{ss}/2]$, respectively, where $i=1,2$, and $j=1,2$.

According to Eqs. (4) and (5), the optimal spacing range of the system is $0.39\lambda \leqslant r_{ss} \leqslant 1.12\lambda$. Two spacings of the control sources, one $r_{ss}=0.6\lambda$ within the optimal range and other $r_{ss}=1.2\lambda$ outside the range, were chosen to examine the ability of the control system in creating quiet zones. The sound pressure attenuations were measured both in the $x-y$ plane and the $z-y$ plane in the area around and behind the error microphone grid. Figure 7 describes the measured sound attenuation in an $x-y$ plane, which agrees with the result of numerical analysis.[4]

When the spacing of the control sources is chosen to be within the optimal range, the control system can create a large quiet zone in the area behind the error microphones, as shown in Fig. 7(a). If the spacing of the control sources is longer than the upper limit of the optimal range, not only does the system become useless in attenuating the noise, but it causes a large increase of sound pressure everywhere. Figure 7(b) shows that, in some places, this increase can be as high as 10 dB.

IV. SUMMARY

Similar to the case of control sources and error microphones in two parallel lines, the control system with multiple control sources and error microphones evenly placed in two parallel planes also has an optimal range of spacing among the adjacent control sources and error microphones. When the system is arranged within this optimal range, the system can create a large area of quiet zone in free space and the total power output of the control system bears the least increase.

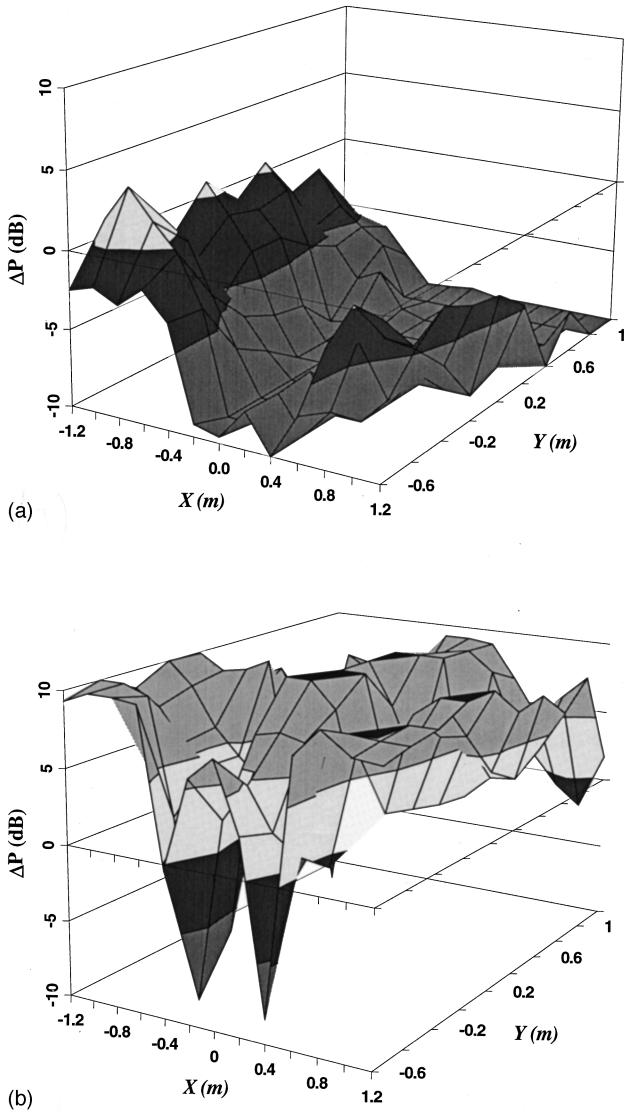


FIG. 7. Measured sound pressure attenuation in an anechoic chamber with the spacings of the control sources (a) $r_s=0.6\lambda$ and (b) $r_s=1.2\lambda$.

[1] J. Guo and J. Pan, ''Analysis of active noise control in a free field,'' Proceedings of Active 95, Newport Beach, CA, pp. 649–660 (1995).
[2] J. Guo, J. Pan, and C. Bao, ''Actively created quiet zones by multiple control sources in free space,'' J. Acoust. Soc. Am. **101**, 1492–1501 (1997).
[3] S. E. Wright and B. Wuksanovic, ''Active control of environmental noise,'' J. Sound Vib. **190**, 565–585 (1996).
[4] J. Guo, ''Active noise control in open space and active noise barriers,'' Ph.D. thesis, University of Western Australia, 1997.

What should be the goals of cochlear modeling?

R. S. Chadwick
LCB, NIDCD, National Institutes of Health, 9 Center Dr. MSC 0922, Bethesda, Maryland 20892-0922

(Received 28 April 1997; revised 26 June 1997; accepted 27 June 1997)

The recent thought-provoking paper by de Boer, ''Classical and nonclassical models of the cochlea,'' not only deserves comment on its own right, but more generally it should cause cochlea modelers to reflect on their ultimate goals. While his claim that a ''classical'' model can be found that has the same response as a ''nonclassical'' model is mathematically correct for linear, single degree-of-freedom models, this paper argues that his algorithm produces a ''nonrealizable'' and hence less useful model. Also, his assertion of uniqueness needs to be qualified.
[S0001-4966(97)03511-X]

PACS numbers: 43.64.Kc, 43.64.Bt [RDF]

The goals of cochlear modeling depend on both the scientific environment of the modeler and how the work is ultimately assessed by others. From my perspective, the purpose of cochlear modeling is to help understand the physical and biological mechanisms involved in the hearing process. If a model can be found that reproduces measured responses with measurable physical and biological parameters, then an important step will have been taken in furthering this understanding. Most cochlea modelers would probably agree that such a model has yet to be discovered. Models such as this are the most useful, and we might refer to them as ''realizable models.'' Models defined by measurable parameters, but which do not reproduce measurable responses are in a sense still useful, since they at least tell us that the cochlea does not work in the way they suggest. Models that reproduce measured responses, but with nonmeasurable physical and biological parameters cannot further our understanding of the hearing process and are therefore less useful. Models such as this we might call ''nonrealizable models.''

I claim that de Boer's algorithm (de Boer, 1997) would produce a ''nonrealizable'' model from a potentially ''realizable'' one. If the *effective basilar membrane impedance* is computed by solving a ''nonclassical'' model and it is used as the *local impedance* in a ''classical'' model as he suggests, the resulting distribution of physical and biological parameters may not be realizable. de Boer recognized this by saying that the *effective basilar membrane impedance* may well be *noncausal*. de Boer's procedure requires several steps: (1) fixing the frequency; (2) computing the solution of the ''nonclassical'' model at that frequency; (3) forming the ratio of pressure and partition velocity; (4) identifying the spatial distribution of physical and biological parameters in a ''classical'' model that results in the same ratio as determined in (3). When this procedure is repeated at a different input frequency, the resulting spatial distributions of physical and biological parameters in a ''classical'' model will be different. Thus in order for a ''classical'' model to reproduce a mechanical tuning curve or transfer function of a ''nonclassical'' model, the mass, stiffness, etc., in the ''classical'' model would have to alter themselves at a given location when the frequency is varied. This property implies not only a loss of uniqueness, but also results in a ''nonrealizable'' and hence less useful model.

A current focus of cochlear mechanics is to achieve a better understanding of how outer hair cell motility contributes to amplification and sharpening of waves on the cochlear partition. It is very unlikely that experimentation or modeling alone can provide definitive answers. Cochlea modelers can increase their contribution to this field if their models raise new questions that suggest new experiments. This type of interaction is facilitated by the use of ''realizable models.''

ACKNOWLEDGMENTS

The author is grateful for many useful discussions with Kuni Iwasa and Emilios Dimitriadis.

de Boer, E. (**1997**). ''Classical and nonclassical models of the cochlea,'' J. Acoust. Soc. Am. **101**, 2148–2150.

Does nonlinear capacitance of the outer hair cell really improve its high-frequency response?

K. H. Iwasa and R. S. Chadwick
Biophysics Section, LCB, NIDCD, National Institutes of Health, 9 Center Drive, MSC 0922, Bethesda, Maryland 20892-0922

(Received 17 April 1997; revised 20 June 1997; accepted 21 June 1997)

There is a conceptual error in the paper, ''A theoretical basis for the high-frequency performance of the outer hair cell's receptor potential'' by El S. Hassan [J. Acoust. Soc. Am. **101**, 2129–2134 (1997)]. Correcting the error reverses the conclusion, and reveals that the nonlinear capacitance in the outer hair cell is a factor that degrades, rather than improves, the high-frequency response of the receptor potential of the outer hair cell. © *1997 Acoustical Society of America.* [S0001-4966(97)02011-0]

PACS numbers: 43.64.Ld, 43.64.Kc, 43.64.Bt [RDF]

The outer hair cell (OHC) is a mechanoreceptor cell with a voltage-dependent motility (Brownell *et al.*, 1985). This characteristic of the OHC, together with its strategic location between the basilar membrane and the reticular lamina, makes the cell very important for micromechanics of the cochlea. It has been shown that the motor function of the cell is responsive to the membrane potential throughout the auditory frequencies, and yet the receptor potential of the OHC has a low-pass characteristic with estimated roll-off frequencies less than 1 kHz, limiting the effectiveness of the OHC in responding to the vibration of the organ of Corti cycle-by-cycle (Housley and Ashmore, 1992; Dallos and Evans, 1995).

The paper by El S. Hassan (1997) addresses this question. However, the author made a conceptual error in defining a nonlinear capacitance, and introduced an unphysical term. Capacitance C is usually a constant defined as the ratio of the charge Q to the voltage V. In the general case where capacitance is not necessarily constant, it is naturally extended as a coefficient relating an infinitesimal charge increment dQ with an infinitesimal voltage increment dV, i.e., $dQ=C\,dV$. In other words, the charge Q is represented by $\int C\,dV$. The nonlinear capacitance in the outer hair cell has been obtained based on this definition (Santos-Sacchi, 1991; Iwasa, 1993). Hassan, however, defines the capacitance C by $Q=CV$, even though the capacitance C is not constant. This leads to $dQ=CdV+VdC$, giving rise to an additional term. In addition, he confuses his capacitance with the one which is defined differently. He thus introduces an unphysical term, which leads to a false reduction in the RC time constant. Based on this result, he interprets that the nonlinear capacitance is a mechanism for extending the magnitude of the receptor potential to higher frequencies.

To illustrate the point, let us consider an outer hair cell in the whole-cell recording configuration in which the access resistance is small. In addition, the reversal potential of the conductive element is omitted for simplicity because it is not essential for discussing AC currents. Then the system consists of parallel connection of a conductance g_h, a linear capacitance C_h, and a nonlinear capacitance C_u. Suppose a voltage V is applied to this circuit. In Hassan's interpretation, the current I would be written in a manner similar to the basolateral current of his first equation [the right-hand side of his Eq. (1)],

$$I=g_hV+C_h\frac{dV}{dt}+\frac{d}{dt}(C_uV)$$

$$=g_hV+\left(C_h+C_u+V\frac{dC_u}{dV}\right)\frac{dV}{dt}, \tag{1}$$

with a nonlinear capacitance (Ashmore, 1990; Santos-Sacchi, 1991; Iwasa, 1993)

$$C_u(V)=\frac{2C_{\max}}{1+\cosh[q(V-V_p)/k_BT]}, \tag{2}$$

where q is the gating charge of the motor, V_p the voltage at peak capacitance, k_B the Boltzmann constant, and T the temperature. If we put $q=-e$ (e: electronic charge) and $V_p=-30$ mV, the term $V\cdot dC_u/dV$ is negative in the region $V<-56$ mV, with the minimum at around -90 mV. For this reason, the total capacitative current can be nullified, as shown in Fig. 3(a) (Hassan, 1997). That is the source of the reduced RC time constant discussed in the paper.

Let us examine the origin of the nonlinear capacitance as expressed by Eq. (2). The charge function of the two-state membrane motor can be expressed (Iwasa, 1994),

$$Q_u(V)=Q_1P_1+Q_2(1-P_1), \tag{3}$$

where Q_1 and Q_2 are the charges in the two states of the motor, and the probability P_1 that the motor is in one of the states follows the Boltzmann distribution,

$$P_1=\frac{1}{1+\exp[q(V-V_p)/k_BT]}.$$

The motor obtains energy by transferring its charge across the membrane. The electrical capacitance $C_u(V)$ due to the motor is obtained as the voltage derivative of the membrane charge $Q_u(V)$,

$$C_u(V)=\frac{dQ_u(V)}{dV}. \tag{4}$$

The combination of Eqs. (3) and (4) leads to Eq. (2). Equation (3) was also used for experimental determination of the nonlinear capacitance (Santos-Sacchi, 1991). Strictly speaking, the form of the nonlinear capacitance given is in the low-frequency limit. It rolls off at the frequency at which the motor charge cannot follow voltage changes. This roll-off frequency is related to the response time of the motor. Now the capacitative current due to the motor charge is given by

$$\frac{dQ_u(V)}{dt} = \frac{dQ_u(V)}{dV} \cdot \frac{dV}{dt} = C_u(V) \frac{dV}{dt}.$$

Thus the whole cell current should be

$$I = g_h V + [C_h + C_u(V)] \frac{dV}{dt}. \tag{5}$$

Equation (5) disagrees with Eq. (1). The effect of the nonlinear capacitance on the circuit property of the hair cell is to make the RC time constant larger rather than smaller because $C_u(V)$ is always positive. Equation (5) is consistent with Eq. (2), not only theoretically but experimentally, because the nonlinear capacitance C_u as expressed by Eq. (2) was also determined by measuring membrane currents based on Eq. (5) (Santos-Sacchi, 1991; Iwasa, 1993). This means that Eq. (1) is incorrect and incompatible with Eq. (2).

In fact, the relationship $Q(V) = C(V)V$, as assumed in Eq. (1), is valid only if $C(V) = \text{const}$. This error led to an extra term $V \cdot dC_u/dV \cdot dV/dt$, which reduced the capacitative current in Hassan's calculation, as it is illustrated earlier.

The above argument is based on the definition of the capacitance given by Eq. (4). In Hassan's definition, the capacitance would be $(Q_u(V) - Q_u(0))/V$ instead of Eq. (2). The subtraction with the charge at zero potential $Q_u(0)$ is to avoid divergence at $V = 0$. Mathematically, the source of the mistake in Hassan's paper is in confusing $(Q_u(V) - Q_u(0))/V$ with dQ_u/dV. Hassan's capacitance, however, is physically not a natural extension of the regular capacitance. It is not useful in describing a system. For example, if the quantity $(Q_u(V) - Q_u(0))/V$ is used instead of Eq. (2) for plotting experimental data, a voltage shift in $Q_u(V)$ cannot be easily recognized.

In conclusion, the nonlinear capacitance contributes to the RC time constant, together with the regular capacitance. Thus a factor which makes the motility of the outer hair cell effective on a cycle-to-cycle basis must be sought elsewhere.

Ashmore, J. F. (**1990**). ''Forward and reverse transduction in the mammalian cochlea,'' Neurosci. Res. Suppl. **12**, S39–S50.

Brownell, W. E., Bader, C., Bertrand, D., and Ribaupierre, Y. (**1985**). ''Evoked mechanical responses of isolated cochlear hair cells,'' Science **227**, 194–196.

Dallos, P., and Evans, B. N. (**1995**). ''High-frequency motility of outer hair cells and cochlear amplifier,'' Science **267**, 2006–2009.

Hassan, E.-S. (**1997**). ''A theoretical basis of the high-frequency performance of the outer hair cell's receptor potential,'' J. Acoust. Soc. Am. **101**, 2129–2134.

Housley, G. D., and Ashmore, J. F. (**1992**). ''Ionic currents of outer hair cells isolated from the guinea pig,'' J. Physiol. (London) **448**, 73–98.

Iwasa, K. H. (**1993**). ''Effect of stress on the membrane capacitance of the auditory outer hair cell,'' Biophys. J. **65**, 492–498.

Iwasa, K. H. (**1994**). ''A membrane motor model for the fast motility of the outer hair cell,'' J. Acoust. Soc. Am. **96**, 2216–2224.

Santos-Sacchi, J. (**1991**). ''Reversible inhibition of voltage-dependent outer hair cell motility and capacitance,'' J. Neurosci. **11**, 3096–3110.

PROGRAM OF

The 134th Meeting of the Acoustical Society of America

1–5 December 1997

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, 1 DECEMBER 1997 EL CAMINO/ADOBE ROOMS, 8:15 A.M. TO 12:00 NOON

Session 1aAO

Acoustical Oceanography: Acoustic Measurement of Ocean Processes

Eric Wolin, Chair
Hughes Aircraft, Naval and Maritime Systems, 8080 Dagget Street, San Diego, California 92111

Contributed Papers

8:15

1aAO1. Acoustic monitoring of flow through the Strait of Gibraltar. Peter F. Worcester (Scripps Inst. of Oceanogr., Univ. of California at San Diego, La Jolla, CA 92093, pworcester@ucsd.edu), Uwe Send (Univ. of Kiel, 24105 Kiel, Germany), Bruce D. Cornuelle, and Christopher O. Tiemann (Univ. of California at San Diego, La Jolla, CA 92093)

The Strait of Gibraltar Acoustic Monitoring Experiment was conducted during April–May 1996 to determine the feasibility of using acoustic methods to make routine, rapidly repeated, transport measurements in the Strait of Gibraltar, as well as to explore the acoustic scattering caused by the internal wave bores generated in the Strait. Three different approaches to monitoring the flow were explored: (i) high-frequency (2-kHz) reciprocal transmissions; (ii) high-frequency (2-kHz) horizontal arrival angle measurements of ray bending due to currents; and (iii) one-way transmissions from a low-frequency (250-Hz) source to a vertical receiving array on the opposite side of the Strait. Extensive independent measurements of the temperature, salinity, and velocity fields were also made. The deep-turning rays are stable and yield differential travel times that give reasonable values for the current components along the acoustic paths. Horizontal arrival angles (phase differences) are strongly correlated with rough estimates of the current perpendicular to the acoustic path made using current meter data, suggesting that this approach is feasible at ranges up to at least 15 km. Finally, the acoustic signals show the effects of the nonlinear internal waves present in the Strait. [Work supported by ONR.]

8:30

1aAO2. Direction and speed of internal waves in SESAME II, and their effect on acoustic transmission. Jacob George, Robert L. Field, and Zachariah R. Hallock (NRL, Stennis Space Center, MS 39529)

Analysis of thermal data from SESAME II has revealed variations of both long and short durations. Variations of long durations have been clearly identified with tidal cycles, and are nearly identical at two thermal arrays 433 m apart in range, with no significant relative time lag between the two. Though a zero relative time lag would imply a southeastern direction for the wave vector, the long wavelengths involved and the short range between the arrays make such a prediction inconclusive. Variations of shorter durations arrive at the two arrays with a relative time lag that ranges between 5 and 26 min. Wavelet analysis has confirmed this time lag, and shows that it holds for a range of wavelengths. This puts constraints on the direction and speed of the short-duration components. The implications of these results on acoustic propagation will be discussed. [Work supported by ONR and NRL.]

8:45

1aAO3. Forward propagation analysis for the summer Shelfbreak PRIMER Experiment. Brian J. Sperry, Jim Lynch (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), Ching-Sang Chiu (Naval Postgrad. School, Monterey, CA), Glen G. Gawarkiewicz, Robert S. Pickart (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), James H. Miller (Univ. of Rhode Island, Narragansett, RI), and Allan R. Robinson (Harvard Univ., Cambridge, MA)

One of the primary objectives of the 1996 Shelfbreak Primer Experiment, located in the Mid-Atlantic Bight south of Nanucket Shoals, was to investigate the effects of the shelfbreak front on the propagation of acoustic signals. The acoustic component of the experiment consisted of two 16-element vertical hydrophone arrays (VLAs) along the northern edge of the region (inshore of the shelfbreak), and three sources (one 224 Hz and two 400 Hz) along the southern edge (seaward of the shelfbreak). Source-to-receiver path lengths ranged from 40–60 km. The high-quality oceanographic data collected, including SeaSoar sections and numerous thermistor records, are critical to modeling and interpreting the acoustics. From a modal perspective, there is strong coupling from not only the front, but also from the sloping bathymetry and from scattering by solitons. The receptions at the northeast VLA coinciding with a 6-day period of SeaSoar

deployment are considered here. Fluctuations in the signal arrivals are consistent with the picture provided by the measured oceanography. Results from coupled-mode propagation modeling are also discussed. [Work supported by ONR.]

9:00

1aAO4. Pulse propagation through continental-shelf internal solitary waves. Altan Turgut and Stephen N. Wolf (Naval Res. Lab., Acoust. Div., Washington, DC 20375)

A newly developed pseudospectral numerical model is used to study pulse propagation through continental-shelf internal waves. A time-domain mode coupling analysis is performed on plane-wave pulses propagating through a simple internal solitary wave (ISW) described by downward undulation of a thermocline having the shape of hyperbolic secant. Initial acoustic fields made up of eight-cycle Hanning-weighted sine waves with 224- and 400-Hz center frequencies (vertically shaded by desired acoustic mode shapes) are used. Two-dimensional and three-dimensional numerical results were compared to investigate the validity of two-dimensional techniques in the case of nonparallel acoustic-ISW alignment. Similar to the results of Preisig and Duda [IEEE J. Oceanic Eng. **22**(2) (1997)], it has been observed that the horizontal refraction and azimuthal scattering seem to be minor for up to 65-deg oblique incidence angle. Results of two-dimensional propagation and mode-coupling analysis for the ISW packets measured during the SWARM95 experiment are also discussed. [Work supported by ONR.]

9:15

1aAO5. Acoustic field propagation through a shallow water waveguide dominated by internal waves. Steven Finette, Marshall Orr (Acoust. Div. Naval Res. Lab., Washington, DC 20375), John Apel (Global Ocean Assoc., Silver Spring, MD 20908), and The SWARM Group

The continental shelf off the New Jersey coast is a site of intense internal-wave activity. The space–time variability of an acoustic field in this shallow-water region was studied using oceanographic measurements along a 42-km track from the SWARM95 [shallow-water acoustics in a random medium] experiment. A sound–speed model based on the KdV equation was developed to include the nonlinear contribution of the internal wave field, with parameters estimated from towed yo-yo ctd data, ship radar images of internal-wavesurface expression and acoustic backscatter images of subsurface internal-wave structure. The linear internal-wave contribution was modeled by an ADCP-derived displacement power spectrum. Bathymetry and sub-bottom parameters were previously estimated from wave-field inversions using chirp sonar surveys. Simulation results are presented for acoustic propagation at frequencies of 224 and 400 Hz using a wide-angle parabolic equation method to compute realizations of the acoustic field. Results include estimates of the intensity fluctuations and coherence for two vertical arrays located at 33 and 42 km from the moored sources. [Work supported by ONR.]

9:30

1aAO6. Normal mode analysis of acoustic thermometry of ocean climate receptions using short-time Fourier techniques. Kathleen E. Wage, Arthur B. Baggeroer (Res. Lab. of Electron., MIT, Cambridge, MA 02139 and Appl. Ocean Phys. and Eng. Dept., Woods Hole Oceanogr. Inst.), and James C. Preisig (Woods Hole Oceanograph. Inst.)

The effects of internal waves on the coherence of propagating modes and the validity of the adiabatic assumption are very important issues in both matched-field processing and acoustic tomography. While much theoretical research has been done on long-range propagation of modes in deep water, there have been few opportunities to compare theoretical predictions with experimental measurements. Deployment of mode-resolving vertical arrays as a part of the Acoustic Thermometry of Ocean Climate (ATOC) experiment has provided data sets, which may offer insights about mode coherence at long ranges. This study examines the modal composition of ATOC receptions using the short-time Fourier techniques described in earlier work [Wage *et al.*, J. Acoust. Soc. Am. **101**, 3113(A) (1997)]. Initial results reveal a frequency selectivity in the arrivals for the lowest ten modes. Since frequency-selective effects can be modeled by a random multipath channel, the problem of estimating channel characteristics from pulse-compressed M-sequence receptions is discussed. In particular, this study investigates various techniques of averaging across receptions to obtain the mean arrival time and spread for each mode. [Work supported by ONR.]

9:45

1aAO7. MFP geoacoustic inversion applied to the WORKSHOP97 data. A. Tolstoy (Integrated Performance Decisions, Inc., Honolulu, HI 96816)

Application of the refined iterated grid search (RIGS) MFP inversion method to four of the WORKSHOP97 test cases (SD, AT, SO, and WA) has resulted in some surprisingly accurate estimates for some geoacoustic parameters, i.e., sediment densities and attenuations, source ranges and depths, and some sediment gradients. Another surprise was the sensitivity of the inversion to small errors in the KRAKEN propagation model. This paper will discuss recent efforts to understand the benefits and shortcomings of the RIGS method and to propose future directions.

10:00–10:15 Break

10:15

1aAO8. Modeling of broadband time signals in shallow water using environmental inversion. Peter L. Nielsen (Saclantcen, Viale S. Bartolomeo 400, 19138 La Spezia, Italy, nielsen@saclantc.nato.int), Peter Gerstoft (UCSD, La Jolla, CA 92093-0704), Francesco Bini-Verona, and Finn B. Jensen (Saclantcen, La Spezia, Italy)

A numerical modeling scheme is applied to perform broadband environmental inversion by using explosive data acquired in the Mediterranean Sea in May 1997. The area shows a moderate range dependency, and the received time signals from explosive charges cover a frequency band of several kilohertz. The acoustic propagation model is based on a layered normal mode approach, which is capable of efficiently predicting broadband sound propagation in shallow water [Westwood *et al.*, ''A normal mode model for acoustic–elastic ocean environments,'' J. Acoust. Soc. Am. **100**, 3631–3645 (1996)]. This model has been extended to handle range-dependent environments using the adiabatic approximation. Thus calculation of broadband transfer functions from 0 to 10 kHz of range-dependent, shallow-water waveguides can now be done within a few minutes. The combination of the above propagation model and a state-of-the-art global inversion scheme [Gerstoft, ''SAGA user manual 2.0: An inversion software package,'' SACLANTCEN SM-333 (1997)], makes it possible to perform environmental focusing by optimizing the correlation between the numerical and experimental received time signal at one or more receiver locations. The broadband inversion scheme is introduced to extract the uncertain acoustic parameters in the waveguide, and to assess how accurate broadband signals can be modeled in complex shallow-water regions.

10:30

1aAO9. Broadband mode inversion for oceanographic and geoacoustic parameters. Gopu R. Potty and James H. Miller (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882)

A genetic algorithm was used for the inversion of sound speed in the water column and sediment layers south of New England in the Middle Atlantic Bight. The experimental data were SUS charge explosions acquired on a vertical hydrophone array during the Shelf Break Primer Experiment conducted in August, 1996. A genetic algorithm was used to search for optimal parameters in a wide space; the EOF coefficients in the ocean and the compressional sound–speed values in the sediment layers. A range-independent normal-mode routine was used to construct the replica fields corresponding to the parameters. Comparison of group speeds for modes 1–4 and for a range of frequencies from 10–100 Hz was used to arrive at the best parameter fit. A linear perturbative inversion also was

done on the same data as described by Rajan *et al.* [J. Acoust. Soc. Am. **82**, 998–1017 (1987)]. This linear method also gives estimates of resolution and variance. Analysis was also done to compute the degree of convergence of each of the parameters. Group speeds for the inverted sound–speed fields provide an excellent match to the experimental data. An efficient hybrid optimization scheme using genetic and linear techniques is presented. [Work supported by ONR.]

10:45

1aAO10. Observations of attenuation during the Scripps Pier bubble experiment. Jerald W. Caruthers, Paul A. Elmore, Philip A. Beben, and Stephen J. Stanic (Naval Res. Lab., Ocean Acoust. Branch, Stennis Space Center, MS 39529)

An experiment that measured the effects of bubbles just offshore from active surf was performed in the spring of 1997 off the pier at the Scripps Institution of Oceanography. A region from a few tens of meters to 1000 m from the beach was instrumented by several researchers. An emphasis was placed on a region approximately 400 m^2 in area and about 300-m offshore where most of the instruments were clustered. Located in this area was a triangular-shaped frame called the Delta Frame, which had sources at two vertices and eight hydrophones along the perimeter (each side was 9.4 m long). Eight frequencies, from 39 to 244 kHz, were propagated sequentially from each source and received on each hydrophone. All 16 transmissions occurred within 12 ms and were repeated at a 1-s interval during each of ten 88-min data runs. Bubbles produced in the surf zone were carried out to the experimental region by rip currents and had dramatic effects on all the instruments. This paper discusses their effects on attenuation as measured at the Delta Frame. [Work supported by ONR.]

11:00

1aAO11. Bubble clouds, velocity distributions, and high-frequency propagation in the Scripps bubble experiment. David M. Farmer, Svein Vagle (Inst. of Ocean Sci., 9860 West Saanich Rd., Sidney, BC V8L 4B2, Canada), and Grant Deane (Scripps Inst. of Oceanogr., La Jolla, CA 92093)

As part of an experiment to determine the distribution and acoustical effects of bubbles near the surf zone, resonators, cameras, a coherent Doppler, and horizontally oriented Doppler sonars were deployed in the vicinity of Scripps Pier. With these instruments, evolution of the waves as they traveled toward the shore were detected and the bubble size distributions, rip currents, turbulence dissipation, and patterns of bubble cloud motion were measured. Propagation tests at 100 kHz along a 230-m path between the end of the pier and the surf were conducted. Bubble size measurements in the surf and on a bottom-mounted frame allowed identification of the spatial evolution of bubble characteristics. The sonar measurements show bubble clouds ejected offshore at discrete intervals and at speeds of $\sim$0.3 m/s. Pulse propagation into the surf was occasionally completely attenuated by bubble clouds. However, it is interesting that prior to total blocking of the signal the pulse arrival time is delayed, implying an apparent decrease in sound speed. This is unexpected since our frequency (100 kHz) was above the resonance of the dominant bubbles. An explanation is proposed in terms of selective blocking of direct path propagation by bubble clouds. [Work supported by ONR Contracts Nos. N00014-96-C-6030 and N00014-97-1-0288.]

11:15

1aAO12. Tomographic reconstruction of shallow-water bubble fields observed in the Scripps Pier bubble experiment. Daniel Rouseff, Frank S. Henyey (Appl. Phys. Lab., College of Ocean and Fishery Sci., Univ. of Washington, Seattle, WA 98105), Jerald W. Caruthers, and Steven J. Stanic (Naval Res. Lab., Stennis Space Center, MS 39520)

An experiment was performed in the Spring of 1997 off the pier at the Scripps Institute of Oceanography. The objective was to measure the nearshore properties of bubble fields. A major component of the experiment was the Delta frame designed at NRL-SSC. The triangular frame, 9.4 m on each side, supported two sources and eight hydrophones. Measurements of acoustic travel time and attenuation were made at eight frequencies between 39 and 244 kHz. Details of the measurement apparatus will be presented elsewhere at this meeting [J. W. Caruthers *et al.*, ''Observations of attenuation during the Scripps Pier bubble experiment'']. In this paper, a tomography algorithm [D. Rouseff and F. S. Henyey, J. Acoust. Soc. Am. **101**, 3032 (1997)] is applied to a sampling of the data to produce cross-sectional mappings of attenuation. The bubble concentrations are shown to be inhomogeneous and carried across the frame by rip currents. Possible improvements to the tomography algorithm for imaging strongly attenuating structures are discussed. [Work supported by ONR.]

11:30

1aAO13. Attenuation estimates for bubbly sea water. R. R. Goodman (Appl. Phys. Lab., Penn State Univ., State College, PA), P. A. Elmore, and J. W. Caruthers (Naval Res. Lab., Stennis Space Center, MS 39529)

Acoustic techniques used for determining bubble distributions can be limited because a solution to an ill-conditioned inverse problem is required. Examining cases in which closed-form solutions are possible, however, may give us insight about the inverse procedure. In a previous ASA meeting [Goodman, Caruthers, and Elmore, ''Dispersion estimates for bubbly seawater,'' J. Acoust. Soc. Am. **100**, 3196(A) (1997)], exact solutions for dispersion for the inverse procedure in special cases of bubble distributions were explored. Here some of those special cases are examined in terms of how they apply to attenuation measurements. For cases in which the closed-form solution are not possible, the resonant bubble approximation allows for estimating the bubble population from attenuation measurements. This paper shows that there is observable error in this procedure, but that the error can be reduced by application of an iterative procedure discussed here. [Work supported by ONR.]

11:45

1aAO14. Measuring near surface ocean features using upward facing ADCPs. Brandon S. Strong and R. Lee Gordon (RD Instruments, 9855 Businesspark Ave., San Diego, CA 92131)

This paper presents the results of acoustic measurements of the ocean surface and near-surface features using a conventional broadband acoustic Doppler current profiler (ADCP). An analysis of the trade offs and error sources associated with surface detection reveals how transducer beam geometry and signal processing can affect surface measurement. Configurations are considered for the measurement of mean-surface height, ambient sound, wind speed, wind direction, surface currents, directional wave spectra, tidal flow, and entrained air. Observations of the sea surface conditions at Scripps Pier are compared with acoustic data collected using upward facing ADCPs of different beam geometries and a pressure sensor. Comparisons are made with previous studies [Zedel, Deep ocean wave measurements using a vertically oriented sonar, J. Atmos. Oceanic Tech. **11**, 182–191 (1994) and Terray *et al.*, Measuring wave height and direction using upward-looking ADCPs, to be presented at Oceans '97].

Session 1aPA

Physical Acoustics: Nonlinear Acoustics

Bart Lipkens, Chair
MacroSonics Corporation, 1570 East Parham Road, Richmond, Virginia 23228

Contributed Papers

8:30

1aPA1. Comparison between a nonlinear equation and acoustic measurement in a laminate graphite-epoxy composite. P. A. Elmore[a] and M. A. Breazeale (Jamie Whitten Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

Measurements of longitudinal phase velocity and rate of harmonic generation (nonlinearity parameter) have been made in a laminate graphite-epoxy composite. Both quantities are observed to be frequency dependent between 3.5 and 8 MHz. The theoretical model of Van Den Abeele and Breazeale [J. Acoust. Soc. Am. **99**, 1430–1437 (1996)] has been applied to the data. Including dispersion terms and their nonlinear counterparts improves agreement over that obtained with the theory used for single crystals. The origin of this difference probably is in scattering from laminate boundaries. Some inadequacies still exist, however. Their origin has not been established yet. [a]Current address: Naval Res. Lab., Ocean Acoustics Branch, Stennis Space Center, MS 39529.

8:45

1aPA2. Modeling of harmonic generation and shock formation in nonlinear surface acoustic waves in several real crystals. R. E. Kumon, M. F. Hamilton (Dept. of Mech. Eng., Univ. of Texas, Austin, TX 78712-1063), Yu. A. Il'inskii, and E. A. Zabolotskaya (MacroSonix Corp., Richmond, VA 23228)

Harmonic generation and shock formation in nonlinear surface acoustic waves that propagate in anisotropic crystals were studied numerically on the basis of a new theoretical model presented earlier [Hamilton *et al.*, *Nonlinear Acoustics in Perspective*, edited by R. J. Wei (Nanjing University Press, Nanjing, 1996), pp. 64–69]. The theory applies for arbitrary elastic materials, surface cuts, and propagation directions. Numerical simulations were performed for initially sinusoidal signals propagating across the surface of KCl, Ni, and Si crystals for the (001), (110), and (111) cuts, and over the appropriate range of directions for each cut. Waveforms are shown to exhibit asymmetric distortion, well-defined shocks with cusped spikes in the horizontal waveform, and a phase shift in the zero crossings. Solutions for propagation in the (001) plane and in the ⟨100⟩ direction of KCl are shown to exhibit atypical trapping of energy in the lowest order harmonics. Analytical solutions derived for the fundamental and second-harmonic components for this particular case are in good agreement with the numerical solutions close to the source. Although the phenomenon resembles one observed in nonlinear optics resulting from dispersion, it is due instead to properties of the nonlinearity coefficient matrix. [Work supported by ONR.]

9:00

1aPA3. Measurements of macrosonic standing waves in oscillating cavities. Christopher C. Lawrenson, Bart Lipkens, Timothy S. Lucas, David K. Perkins, and Thomas W. Van Doren (MacroSonix Corp., 1570 East Parham Rd., Richmond, VA 23228)

Measurements of macrosonic standing waves in oscillating closed cavities are shown. These cavities (resonators) were designed by MacroSonix using resonant macrosonic synthesis (RMS) to shape the resultant waveform. By controlling the nonlinear processes by which energy is transferred to harmonic frequencies, RMS allows design of resonators that give high-amplitude shock-free waveforms. Measurements in cavities designed with RMS show standing-wave overpressures in excess of 340% of ambient pressure, compared to maximum overpressures in cylindrical cavities of about 17%. Power is delivered by oscillating the entire resonator along its axis with a linear actuator (entire resonator drive). Measurements are shown for four axisymmetric resonator shapes: cylinder, cone, horn-cone hybrid, and bulb. Resonators were filled with nitrogen, propane, or refrigerant R-134a (1,1,1,2-tetrafluoroethane). Ratios of peak-to-minimum pressures of 27 were observed. Since practical compressors for air, refrigerants, or other gases require pressure ratios (discharge to suction) of 3 or more, RMS technology can be used in a wide range of applications. Frequency sweeps show softening or hardening behavior, depending on resonator shape. High-amplitude resonance sweeps show significant hysteresis.

9:15

1aPA4. A theoretical model of nonlinear standing waves in an oscillating cavity. Yurii A. Ilinskii, Bart Lipkens, Timothy S. Lucas, Thomas W. Van Doren, and Evgenia A. Zabolotskaya (MacroSonix, 1570 East Parham Rd., Richmond, VA 23228)

A theoretical investigation of nonlinear standing waves in an acoustical resonator is presented. The motivation for this research stems from the new technology of resonant macrosonic synthesis (RMS) developed at MacroSonix. RMS creates high amplitude standing waves, e.g., overpressures in excess of 300% of ambient pressure. The analysis is based on a one-dimensional model equation for the velocity potential that is derived from the fundamental gas dynamics equations for an ideal gas. Nonlinearity, gas viscosity, and entire resonator driving are included. The resonator is assumed to be of an axisymmetric, but otherwise arbitrary, shape. Nonlinear spectral equations are integrated numerically for a two-point boundary-value problem. The harmonic amplitudes and phases of the velocity potential wave are obtained directly from the solution of the frequency-domain equations. The pressure wave shape, the harmonic amplitudes of the pressure wave, and harmonic amplitude distribution along the resonator axis are then calculated. Results are presented for three resonator geometries: a cylinder, a cone, and a bulb. Both hardening and softening behaviors are observed and shown to be geometry dependent. At high amplitude, hysteresis effects are present in the frequency-response curves. Comparisons between measured and calculated waveforms show good agreement.

9:30

1aPA5. Variation of B/A with $1/c$ for several liquids. M. Paul Hagelberg (Dept. of Phys., Wittenberg Univ., P.O. Box 720, Springfield, OH 45501-0720)

Theoretical models [B. Hartmann, J. Acoust. Soc. Am. **65**, 1392–1396 (1979), B. Frank and J. D. N. Cheeke, J. Acoust. Soc. Am. **101**, 1184–1186 (1997)] predict that the variation of B/A, the acoustical nonlinearity parameter, with the inverse sound speed $1/c$, should be linear or linear with a quadratic term. Confirmation of these models requires experimental data for B/A vs $1/c$ for individual liquids. Such data will be presented for several liquids including pure water, seawater, mercury, and a number of organic liquids.

9:45–10:00 Break

10:00

1aPA6. An improved theoretical model for highly nonlinear bubbly liquids. Zhemin Zhu, Xiaoliang Zhao, and Gonghuan Du (Inst. of Acoust. and State Key Lab., Nanjing Univ., Nanjing 210093, P. R. China)

Strong nonlinearity of bubbly liquids is a most interesting property not only for its theoretical importance in nonlinear acoustics but also for its potential application in, for example, ultrasound imaging technology. A physical model which describes the nonlinear property of liquids containing uniform size bubbles is presented [Ultrasound Med. Biol. **21**, 545–552 (1995)]. The theory is successful to some extent but still has some limitations since the sizes of the bubbles are usually not exactly the same. In this paper, a model dealing with more practical situations, i.e., liquids containing many bubbles whose sizes are randomly distributed is extended. The effective nonlinearity parameter B/A (as high as about 10^4) and attenuation coefficient for these bubbly liquids are given. Calculations and comparisons show that the improved model explains the existent experiments more satisfactorily. [Work supported by the National Science Foundation of China.]

10:15

1aPA7. Nonlinear phenomena of high-amplitude vibration of a piston in a gas-filled liquid. Oleg Rudenko and Valery Andreev (Dept. of Acoust., Phys. Faculty, Moscow State Univ., Moscow 119899, Russia)

Nonlinear phenomena are known to be significant after the wave of moderate amplitude passes a long distance through a weakly dissipative medium. Such ''accumulative'' effects are studied exhaustively by nonlinear wave physics. In contrast, the ''local'' nonlinearities are studied incompletely. Such nonlinear behavior can be demonstrated by a piston immersed in liquid and vibrating with high amplitude. This piston subjected to a harmonic load can radiate not only the fundamental frequency but high-order harmonics as well. Moreover, reaction to high-power radiation can create an additional nonlinear resistance to a piston motion. Local nonlinear phenomena are expessed clearly at vibration velocities comparable with the sound speed of the surrounding medium. Such a case can be realized using liquids containing gas bubbles where sound velocity is much less than one in a pure liquid. This work is devoted to the theoretical calculation of the temporal and spectral characteristics of a nonlinear wave radiated by a piston subjected to a harmonic external force. Two different problems are discussed corresponding to the piston considered as a linear and as a nonlinear vibrating system. [Work supported by RFFI and CRDF.]

10:30

1aPA8. Shock-wave collisions in bubbly liquids: Numerical studies. Valery K. Kedrinskii (Lavrentyev Inst. of Hydrodynam., Lavrentyev prospect 15, Novosibirsk 630090, Russia)

The results of numerical investigations of different effects arising in bubbly liquid due to wave interactions are presented. The influence of the gas content on the parameters and the structure of shock waves and of chemical reactions in the gas phase will be considered. The problem is solved within the framework of the two-phase mathematical model including the kinetic equations for the description of the medium-state dynamics, thermoexchange as well as of bimolecular reaction kinetics. The purpose is to estimate the amplification level of the wave intensity as the result of such kinds of interactions. The wave amplitudes in a passive bubbly media turned out to be amplified by one order of magnitude in the collision plane and their amplification depends on the bubble concentration k_0 as $p_{\max}=2+24.5\cdot k_0^{1/4}$. The possibilities of simulating the hot-spot mechanism of liquid explosive detonation ignition and the mechanism of the large-scale explosions of containers filled with a fuel by bubbly detonation wave interactions are discussed. [Work supported by RFFR, Grant 96-02-19369.]

10:45

1aPA9. Fast numerical algorithm for simulation of nonlinear acoustic waves with shocks of finite thickness. Oleg A. Sapozhnikov and Vera A. Khokhlova (Dept. of Acoust., Phys. Faculty, Moscow State Univ., Moscow 119899, Russia, olegs@na.phys.msu.su)

In an earlier work [V. A. Khokhlova and O. A. Sapozhnikov, J. Acoust. Soc. Am. **96**, 3321 (1994)], a modified spectral approach was proposed for the description of nonlinear waves containing shocks. An abrupt shock has an analytical high-frequency ω^{-1} asymptote. This asymptotic result was used in the numerical algorithm to model strongly nonlinear waves with a relatively few number of harmonics $N\sim 20$ [Pischal'nikov *et al.*, Acoust. Phys. **42**, 362–367 (1996)]. However, in real dissipative medium the shock front is not a discontinuity, but a transition region of finite thickness. This region can be adequately described by a hyperbolic tangent profile, so that the correspondent wave spectrum at high frequencies is governed by the Fay solution. Here, the Fay spectrum asymptote of the finite thickness shock is used, instead of the ω^{-1} asymptote of the abrupt shock, to derive a set of coupled differential equations for the harmonic amplitudes. Several model problems are considered. It is shown that this method permits increasing the accuracy and stability of the modified spectral approach, and still leads to a reduction in the number of equations by a factor of 10–100 in comparison with direct frequency domain schemes. [Work supported by FIRCA and RFBR.]

11:00

1aPA10. Nonlinear effects for torsional waves in rods with cracklike defects. Igor N. Didenkulov, Alexander E. Ekimov, and Vyacheslav V. Kazakov (Inst. of Appl. Phys., 46 Ulyanov St., Nizhny Novgorod, 603600, Russia)

The enormously high nonlinear response of solids with cracks to an acoustic excitation makes nonlinear methods possible for nondestructive testing (NDT). Among the problems to be solved are the mechanisms of such nonlinear responses and the development of NDT methods. Results of an experimental study of nonlinear effects for elastic waves in metal rods with cracklike defects and their dependence on the type of contacts in the cracks are given. Unlike earlier works concerned mainly with longitudinal and flexural waves in solids, the present experiment shows that a high degree of nonlinearity is found for torsional waves as well. The crack was modeled by cutting a rod and tightly filling a crack with metal plates. Two types of contact were studied: dry and lubricated contacts. Modulation of high-frequency torsional waves (20 kHz and 22.8 kHz) in a rod by low-frequency flexural vibrations was studied. Flexural vibrations were excited in two ways: with a shock and with the help of a vibrator. Amplitude modulation was observed only in a rod with a crack, the level of this modulation drastically decreasing in the presence of liquid lubricant. [Work supported by RFBR, Russia—Grant No. 97-02-17524—and by INCAS, Nizhny Novgorod, Russia.]

Session 1aSA

Structural Acoustics and Vibration: Structural Vibration and Radiation

Courtney B. Burroughs, Chair
Applied Research Laboratory, The Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16801

Contributed Papers

8:15

1aSA1. A vibro-acoustic method for measuring Young's modulus of building materials. Wing Chu (Indoor Environ. Program, Inst. for Res. in Construct., Natl. Res. Council Canada, Ottawa, ON K1A 0R6, Canada)

The current ASTM standards for the measurements of Young's modulus are designed specifically for ceramics, bricks, carbon and graphite materials, and concrete. The techniques used are based on the measurement of either sonic velocity or the resonant frequency in the flexural mode of vibration of the specimen. The latter method has been adapted for determining the Young's modulus of building materials, such as wood, gypsum board, and oriented-strand-board (OSB), in the Acoustics Laboratory of IRC/NRC using modern digital-signal processing. This paper will discuss the method and provide results for a number of building materials tested.

8:30

1aSA2. Radiation of sound by two concentric free-flooded cylindrical shells excited by a point force. K. Steven Kim (Signatures Directorate, Carderock Div., Naval Surface Warfare Ctr., 9500 MacArthur Blvd., West Bethesda, MD 20817-5700)

Radiated acoustic pressure fields are obtained when two concentric free-flooded cylindrical shells of finite lengths are excited by a time-harmonic point force. Integral equations are formulated for the two elastic thin shells with simply supported boundary conditions at both ends. Near and far-field acoustic pressures are calculated and effects of acoustic coupling between two shells are discussed.

8:45

1aSA3. Radiation from small acoustic sources located close to a submerged, compliantly coated cylindrical shell. Michael D. Gray, Gary W. Caille (Undersea Res. Program Office, Georgia Tech. Res. Inst., 505 Tech Way N.W., Atlanta, GA 30318), John R. Bogle, and Peter H. Rogers (Georgia Inst. of Technol., Atlanta, GA 30332)

Measurements were made of the radiation from individual, small acoustic sources located close to a submerged, compliantly coated, thin cylindrical shell in the $1<ka<20$ frequency range, where k is the acoustic wavenumber, and a is the radius of the coated cylinder. The behavior of the radiated pressure field is shown to have two distinct frequency regimes. For $ka>5$, the field is simple, and is well approximated by a source and its negative image. For $ka<4$, the pressure field is characterized by significant spatial variations, owing to contributions from the elastic response of the coated cylinder. Finite-element model results are compared with the measurements, and are used to provide insight into the interactions between the source and the coated shell. [Work supported by ONR.]

9:00

1aSA4. Far-field acoustic holography onto cylindrical surfaces using pressure measured on semicircles. Andrew Norris (Dept. of Mech. and Aerosp. Eng., Rutgers Univ., Piscataway, NJ 08855-0909, norris@jove.rutgers.edu)

A simple formula was recently proposed by Williams [J. Acoust. Soc. Am. **99**, 2022–2032 (1996)] for imaging pressure and velocity on a vibrating circular cylindrical shell using the far-field pressure measured along a meridional semicircle. The method is discussed and some new results are presented. The procedure is generalized to handle cylindrical surfaces of a noncircular but convex cross section. It is demonstrated that Williams' formula predicts a supersonic surface intensity, which gives the same meridional energy flux as the exact radiated far-field pressure. A modification of Williams' formula is suggested, which uses pressure data from several neighboring semicircles, although complete spherical coverage is not required. The modified imaging formula is based upon the first two terms in an asymptotic expansion in the dimensionless wave number. The leading-order term yields the original formula, and the second term results in a boundary layer type of correction in the circumferential direction. Numerical examples compare the exact supersonic acoustic intensity on a cylinder with that from the original and the modified formula. These indicate that the circumferential on-surface resolution is significantly enhanced by combining data from neighboring semicircles, even when the total far-field spherical coverage is small. [Work supported by ONR.]

9:15

1aSA5. Acoustic modeling of oscillatory hydrodynamic loading and fluid–structure interaction. Ashok Gopinath and Gary W. Sweany (Dept. of Mech. Eng., Code ME/Gk, Naval Postgrad. School, Monterey, CA 93943, gopinath@nps.navy.mil)

An experimental study has been conducted of forces on a cylinder in a standing acoustic wave. The cylinder is representative of the leg of an offshore structure or platform, while the acoustic field is representative of the oscillatory wave loading on such a structure. A piston oscillator drive mechanism provides the requisite large acoustic amplitudes in the pressure vessel resonator that houses the test cylinder. The use of a high-pressure gas allows the desired high values of the Reynolds numbers to be achieved. Both in-line drag forces and transverse (lift) forces on the cylinder, along with the phasing information, have been measured with the help of suitably mounted strain gauges, and corroborated with existing data in the literature. Higher harmonic forces and resonant interactions, typical of compliant structures, have also been measured and analyzed. This experimental technique appears to have a promising potential for studying the large amplitude, large Reynolds number regime of hydrodynamic loading and the resulting fluid–structure interactions.

1aSA6. Abstract withdrawn

9:30

1aSA7. Modification of the four-load method to compute source impedance in a pipe system. B. S. Sridhara (Dept. of Eng. Technol. and Industrial Studies, Middle Tennessee State Univ., Murfreesboro, TN 37132)

This paper discusses a new method used to obtain the source impedance from the four-load method. The three nonlinear algebraic equations of the four-load method were modified to give three equations of a circle. At each frequency, the intersections of these circles give the real and imaginary parts of the source impedance. A computer program was developed to calculate these intersections and extract the realistic values. A frequency range from 0 to 800 Hz was considered for a 0.2286-m (9-in.) diam pipe system with a loudspeaker as the sound source. At each frequency, the circles were scanned from 0° to 360° in intervals of 0.5° allowing small percentage errors such as 0.001, 0.005, and 0.01. The computed source impedance values were compared with those obtained from the direct method. The results agree well with the decrease in the percentage error.

9:45

1aSA8. Method for broadband measurement of complex shear modulus and Poisson ratio of viscoelastic materials. L. Sheiba (NRaD, 53560 Hull St., San Diego, CA 92152-5001)

This method is based on the inverse solution to the vibration problem of a finite elastic cylinder with reinforced torsion surfaces by rigid septa or plates. In all cases the polymer layer to be tested is bonded to rigid plates; therefore, the radial displacement is zero at the boundary between the polymer layer and the septum. The admittance matrix Y of the element with abovementioned boundary conditions is constructed within the framework of the hypothesis for planar cross sections. The dimensionless shear wave number $k_t(\omega)h$ and compression wave number $k_\lambda(\omega)h$, which are the unknown variables, are obtained by measuring the transfer matrix of a known structure and as a result of solving the transcendental equation in a designated frequency range. Data are attained over a broadband of frequencies and temperatures without dependence on the time-temperature superposition principle.

10:00

1aSA9. A combined experimental and numerical method for computing sound radiation from underwater vibrating structures. John B. Fahnline, Matthew J. Erickson, Dean E. Capone, Stephen A. Hambric, and Courtney B. Burroughs (Appl. Res. Lab., Penn State Univ., 16 Applied Science Building, University Park, PA 16804)

Underwater measurements of the radiated sound power from a vibrating structure are conventionally performed either by direct measurements in a large body of water or by reverberation measurements in a water tank. Both methods have inherent drawbacks. Direct measurements in a large body of water require a large scale operation and are relatively expensive. Reverberation measurements are only accurate at high modal densities, and thus a very large tank is required to obtain accurate low-frequency predictions. To avoid both difficulties, a hybrid approach has been devised. Experimental measurements of the vibrations of the structure are taken in a water tank using a scanning laser Doppler vibrometer, where the laser beam is shone onto the submerged structure through an optical quality window. The surface vibration data are then used to derive the specified normal velocity for a numerical calculation of the radiated sound power radiated in free-field conditions. The laser allows dense surface meshes to be defined such that the calculations can be performed to relatively high frequencies. To demonstrate the general accuracy of the calculations, radiation efficiencies are computed for submerged unbaffled plates, and the results are compared to reverberation measurements and to fully numerical predictions.

10:15–10:30 Break

10:30

1aSA10. Mechanical and radiated power and radiation efficiency of point driven panels. J. Ertel (Phys. Dept., U.S. Naval Acad., 572 Holloway Rd., Annapolis, MD 21402-5026, jpe@nadn.navy.mil), J. Dickey, and G. Maidanik (David Taylor Res. Ctr., Annapolis, MD)

The radiation and partial radiation efficiencies from point and line driven panels were previously defined and investigated by the authors [J. Acoust. Soc. Am. **98**, 2888(A) (1995) and J. Sound Vib. **144**, 71–86 (1991)]. In this paper, the mechanical power dissipated in a point driven fluid loaded panel is studied and compared to the radiated power. The interdependence of the mechanical and the radiated powers is investigated, and they are related as fractions of the total input power. The dependencies of the mechanical and radiated power on frequency, fluid loading, and mechanical loss are further studied. In the present paper, the mechanical power as well as the radiation efficiency is shown to increase with increased damping in a panel while the radiated power decreases, as it must. These results again show the fallibility of the conclusion that ''a higher radiation efficiency necessarily implies more radiated power.'' The results of computer experiments are cited in numerical examples.

10:45

1aSA11. Interaction between wave-number pairs. Jean-François Ille and Jerry H. Ginsberg (G. W. Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

When acoustic-structure interaction is analyzed according to the wave-number-based version of the surface variational principle (SVP), considerable computational effort is required to generate each coefficient in the quadratic sum forming the variational quantity. Examination of the wet-surface impedance, which represents the spectrum of pressure amplitudes generated by a specified spectrum of surface velocity amplitudes, reveals that many cross-impedance terms are very small. This suggests that some coefficients need not be computed. The paper introduces several *a priori*-criteria for selecting the impedance terms to be omitted, based on the supersonic cutoff wave number. Each scheme is assessed by comparing its predictions to the convergent SVP solution. For the nonsymmetric azimuthal harmonics, the field quantities and the radiated power are well predicted if subsonic waves are ignored. In contrast, for the axisymmetric component, substantial errors (6 dB or more) for radiated power arise, unless a broad spectrum of subsonic waves are included in the formulation. The power is mostly reactive, being associated with an evanescent field, even for the supersonic spectrum. The study shows that small interactions between subsonic waves in the axisymmetric case can result in substantial radiative effects that are not modeled in ray theory analyses.

11:00

1aSA12. Does the Rayleigh–Ritz method for vibratory continuous systems actually converge to the analytical eigensolution? Jerry H. Ginsberg (G. W. Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

The Rayleigh–Ritz method for continuous systems uses an N-term series with unspecified coefficients as the trial function for the Rayleigh ratio. Extremizing that ratio leads to N approximate natural frequencies $\omega_j^{(N)}$, $j=1,...,N$. The upper bound theorem states that the true values obtained by solving the field equations are such that $\omega_j \leqslant \omega_j^{(N)}$. The separation theorem states that adding a single term to the aforementioned series yields new estimates $\omega_j^{(N+1)}$, $j=1,...,N+1$, such that the previous $\omega_j^{(N)}$ fall in the intervals between the new values. Taken together, these theorems constitute a proof that the lower eigensolutions obtained from the Rayleigh–Ritz method should converge to the true eigensolution. How-

ever, both theorems assume infinite precision arithmetic. This paper uses the simple case of a cantilever beam to examine the behavior of the Rayleigh–Ritz method with increasing series length. Natural frequencies derived from different classes of kinematically admissible basis functions, drawn from monomials, trigonometric functions, and Bessel functions are examined relative to the upper bound and separation theorems. Most selections fail to yield properly behaved solutions if the series length is extended beyond 10–12 terms. In some cases the eigenvalue solver fails to find real eigenvalues, or to find the correct number of eigenvalues. Some of the basis function sets permit formulating the Rayleigh ratio functional analytically, rather than by numerical integration, but the results are the same. The failure of the method is shown to stem from ill-conditioning that arises as a consequence of similarity in the appearance of higher-order basis functions.

11:15

1aSA13. An overview of interaction between friction and vibrations. Adnan Akay (Mech. Eng. Dept., Carnegie Mellon Univ., Pittsburgh, PA 15213)

Friction is a function of both the interface properties and dynamic response of a friction pair. The true contact area and thus the total time-dependent friction and normal forces change as a result of the relative motion of the surfaces with respect to each other. Friction in a dynamic system has the dual role of both exciting vibrations as well as dissipating vibratory energy. In this presentation, mechanisms by which friction induces vibrations in dynamic systems are reviewed and the fundamental issues involved in friction force modeling are discussed.

11:30

1aSA14. Effects of laser stroke parameters on quantitative vibration measurements with stroboscopic shearography. Benjamin A. Bard (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804, bab132@psu.edu)

Digital shearography is a full-field speckle interferometric technique similar to electronic holography. Traditionally, shearography and holography have been used for vibration measurement with either time averaged or stroboscopic techniques. Time averaging images sinusoidal motion of an object's surface over many periods of vibration, resulting in a fringe pattern representative of the magnitude of the displacement mode shape. Stroboscopic illumination synchronizes short bursts, or *strokes*, of laser light with extrema of vibration. This essentially freezes motion of the vibrating surface, allowing it to be studied with techniques normally reserved for static deformation. In particular, phase stepping, an established technique for capturing a series of images and calculating the exact displacement at every point, can be applied to vibration measurement. If displacement amplitudes calculated from laser strobing and phase stepping are to be used for measurement of peak vibration response, corrections are necessary to compensate for errors due to the finite duration of the laser strokes as well as their improper timing due to system impedance. In this presentation, effects of stroke duration will be expressed mathematically and trends demonstrated experimentally. Potential solutions for stroke offset correction will be discussed and experimental validation provided. [Work sponsored by the PSU Applied Research Laboratory.]

MONDAY MORNING, 1 DECEMBER 1997 CALIFORNIA ROOM, 11:00 A.M. TO 12:30 P.M.

Session 1aSC

Speech Communication: Workshop: Basic Science at the Intersection of Speech Science and Communication Disorders

Lynne E. Bernstein, Cochair
Spoken Language Processing Laboratory, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

Gary Weismer, Cochair
Waisman Center, University of Wisconsin, 1500 Highland Avenue, Madison, Wisconsin 53705

11:00

1aSC1. Chair's introduction to the workshop on basic science at the intersection of speech science and communication disorders. Lynne E. Bernstein (Spoken Lang. Processes Lab., House Ear Inst., 2100 West Third St., Los Angeles, CA 90057) and Gary Weismer (Univ. of Wisconsin, Madison, WI 53705)

In the study of speech perception and production, there is a history of developing, evaluating, and revising models and theories based on data collected exclusively from talkers and listeners judged to be free of communication disorders. Basic research with clinically relevant disorders is sometimes confused with, and dismissed as, clinical/applied research. In other branches of science, including, for example, psycholinguistics and biology, study of clinical disorders is regularly employed as one of the avenues for building and testing more comprehensive theories and/or models, and is viewed as a major source of knowledge concerning the physiological structures responsible for normal processes. The intent of the workshop format is to provide a forum for discussion of the significance of clinical disorders for the scientific areas of speech perception and production. A group of distinguished scientists has been invited: (1) to describe their own work involving disorders in speech production and/or perception; and (2) to comment on scientific, data analytic, philosophical, and pragmatic issues that are specific to the enterprise of incorporating data from clinical disorders into their work and their fields. The organizers encourage the attendance of scientists with and without research involving clinical disorders, as well as students.

Session 1pAO

Acoustical Oceanography: Acoustic Observations of Ocean Ridge Processes

Christian P. de Moustier, Chair
Marine Physical Laboratory, Scripps Institution of Oceanography, 9500 Gilman Drive, La Jolla, California 92093-0205

Chair's Introduction—1:25

Special Lecture

1:30

1pAO1. Underwater acoustics for seafloor geodesy. Fred Noel Spiess (Marine Physical Lab., SIO, UCSD, 9500 Gilman Dr., La Jolla, CA 92093-0205)

Geodetic data, particularly using electronic distance measuring and space-based techniques, are central to terrestrial studies of crustal deformation. A decade ago, no comparable capabilities existed for use in the Earth's extensive ocean-covered areas. It had been clear for some time, however, that underwater acoustics could play a key role, replacing or extending the electromagnetic methods used in nearly all terrestrial systems. Three classes of centimeter-capable systems have emerged. For short ranges, direct path measurements using high frequencies are preferred. At intermediate ranges (1–10 km) marker transponders can be interrogated from an intermediate vehicle and their relative positions calculated. Beyond about 10 km, inadequate knowledge of sound speed limits purely acoustic systems, but a combined GPS/acoustic method has been devised to cope with the need for data over important longer base lines. Since 1990, several systems have been implemented in the context of geologically relevant problems, particularly in the northeast Pacific. As operating experience has grown and real data are being analyzed, useful initial results are emerging in conjunction with insights into the problems that must be solved when moving toward future developments. [Work reported has been supported by NSF, NASA, NOAA, PGC, and USGS.]

Contributed Papers

2:00

1pAO2. Possible seismic evidence for fluid migration at the eastern Juan de Fuca Ridge flank. L. Zuehlsdorff, V. Spiess, and C. Huebscher (Dept. of Geosciences, Univ. of Bremen, Bremen, 28334, Germany)

Young and permeable crust along oceanic ridges is often characterized by large-scale fluid migration depending on permeability and heat distribution. The eastern flank of the Juan de Fuca Ridge off the west coast of North America is buried under an unusual thick sediment cover and is therefore suitable for detailed studies on fluid exchange processes. In autumn 1996 a high-resolution seismic survey was carried out in the vicinity of ODP Leg 168 drill sites to collect detailed information about sedimentary structures and acoustic anomalies in relation to basement morphology. The seismic equipment was optimized for high lateral and vertical resolution to identify small-scale features from the surface down to the basement as potential migration paths for fluids. The acquired seismic sections exhibit numerous vertical zones with reduced reflection amplitudes which may represent paths for fluid advection. The lateral amplitude changes are visible using seismic sources with different frequency content and are apparently related to basement topography. Observations on parallel profiles indicate a 2-D geometry. Theoretical studies suggest that local porosity changes may be the reason for the acoustic visibility of the proposed advection zones.

2:15

1pAO3. Sonar observations of deep ocean hydrothermal flows. Darrell R. Jackson, Christopher D. Jones, Timothy Wen (Appl. Phys. Lab., College of Ocean and Fishery Sci., Univ. of Washington, Seattle, WA 98105), Peter A. Rona, and Karen G. Bemis (Rutgers Univ., New Brunswick, NJ 08903-0231)

A mechanically scanned sonar operating at 330 kHz has been used to image hydrothermal flows on ocean ridges using two different techniques. Scattering from particulates is used to image smoker plumes [Rona *et al.*, Geophys. Res. Lett. **18**, 2233–2236 (1991)], and scintillation of seafloor backscatter is used to image diffuse flows. Results will be presented from cruises on the East Pacific Rise and the northern Cleft segment of the Juan de Fuca Ridge. Plume images have been analyzed to extract physical parameters relevant to plume theory. Observations of diffuse flow employ cross correlation of ping doublets. Scattering theory is used to relate the correlation levels to the variance of temperature fluctuations. [Work supported by the NOAA National Undersea Research Program through the West Coast Undersea Research Center.]

Invited Paper

2:30

1pAO4. Seismoacoustic recordings of a volcanic event on the Mohns Ridge, 1995. Donna K. Blackman and John A. Orcutt (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0225, dblackman@ucsd.edu)

A swarm of earthquakes occurred on an oceanic spreading center north of Iceland in late 1995 and was recorded by U.S. Navy hydrophone arrays in the Norwegian Sea. About two dozen of these events on the Mohns Ridge were detected by onshore seismic arrays. Analysis of the hydrophone array data shows that 7000 events occurred; the greatest number (40/h) took place in a 3-day period in the middle of the 70-day duration of the swarm. Recorded arrivals include P waves, water-borne T waves, PT pairs, and P waves reflected at the seasurface. Separation in arrival times of P and T waves are used to determine relative locations of events and

their spatial evolution throughout the swarm. The locus of activity shifts by 30–40 km during the swarm but steady migration of activity is not apparent. This suggests that surface breaks during dike injection did not occur or, at least, did not generate T waves, or that the swarm was not associated with a simple dike emplacement along the ridge. The time history of the activity, on the other hand, is quite similar to that seen associated with two known volcanic events on the Juan de Fuca and Gorda ridges.

Contributed Papers

2:50

1pAO5. Earthquake studies using under-ice hydrophone data (''Spinnaker''). Robert A. Sohn, John A. Hildebrand (Scripps Inst. of Oceanogr., 8602 La Jolla Shores Dr., La Jolla, CA 92093-0205, ras@mpl.ucsd.edu), and Barbara J. Sotirin (Naval Command, San Diego, CA 92152)

Preliminary studies indicate that the Spinnaker array, a network of hydrophones currently deployed in the Arctic Ocean, may be used to monitor seismicity of the Arctic Basin, as well as long distance teleseismic arrivals from the Southern Hemisphere. The Arctic Basin is a tectonically complex region that has seen few seismic studies, and the Spinnaker array has the potential to monitor some of its most interesting features, such as the Nansen–Gekkel Ridge (the slowest spreading ridge in the world), in near real time. In addition, teleseismic arrivals at the array may have raypaths that coincide closely with the Earth's spin axis. These data may prove useful to current studies of the crystalline alignment and differential spin of the Earth's inner core.

3:05

1pAO6. Locating ridge seismicity near Ascension Island using hydroacoustic and seismic data. Jeffrey A. Hanson and Holly K. Given (Inst. of Geophys. and Planetary Phys., UCSD, La Jolla, CA 92093)

Recently, there has been a surge of interest in monitoring small-scale, mid-ocean ridge seismicity as made possible by the increasing availability of data from military-run hydrophone arrays. In the past, long-term ridge seismicity has been studied using the global networks of seismic stations. Since the young lithosphere at the ridge is relatively weak, it is not able to support large earthquakes and much of the seismicity associated with active deformation is not detected by the seismic networks. In addition to seismic body waves, mid-ocean ridge earthquakes usually produce a T phase—a water-borne acoustic wave—which travels great distances with little attenuation and is well recorded by hydrophones. This paper examines earthquakes from a section of the Mid-Atlantic Ridge near Ascension Island, using data collected from five U.S. Air Force hydrophones that surround the island and the seismic station ASCN located on the island. By integrating the hydroacoustic T phase and the local body-wave phases, earthquakes can be located down to a magnitude of 2.8 on a segment of the ridge with an accuracy of about 5 km.

3:20

1pAO7. Implications of nonfractal seafloor stochasticity on acoustical scattering from the Mid-Atlantic Ridge. Vincent Lupien and Arthur B. Baggeroer (MIT, Cambridge, MA 02139)

Acoustical and bathymetric data were collected near the Mid-Atlantic Ridge as part of the Acoustical Reverberation Special Research Program (ARSRP) in 1993. The wideband time-domain envelope statistics of backscatter from prominent bathymetric features exhibit a non-Rayleigh character which previous researchers have described as event-like. The envelope probability density functions (pdf's) show enhanced tails at high levels which can be a source of active sonar clutter. It is proposed that the origin of such clutter is single-scale, nonfractal roughness. High-resolution bathymetry reveals a power spectral density (PSD) with power-law form. Wavelet analyses reveal a single-scale character to the roughness. In order to test the hypothesis, multiscale and single-scale realizations of the measured PSDs are used in numerical simulations of time-domain backscatter. At the range and azimuthal resolutions in the experiment, the envelope pdfs for multiscale surfaces are Rayleigh, while the single-scale surfaces lead to enhanced tails as observed in the data. The conclusions drawn are that along with an interface's rms height, correlation length, and power spectral shape, one should also be concerned with its *scale structure* because it plays an important role in the physics of wave interaction at random interfaces.

3:35

1pAO8. Fractals, wavelets, and stochastic interface modeling. Vincent Lupien (MIT, Cambridge, MA 02139)

The seafloor is one of many natural interfaces which, when viewed as a random process, exhibits a power law decay, i.e., the power spectral density (PSD) decays as A/f^b, $A>0$, $b>0$ as $f\rightarrow\infty$. Power law spectra are usually taken as evidence that the stochastic process is multiscale or fractal. Multiscale interfaces exhibit features at all scales and their acoustical properties are distinct from those of single-scale interfaces, which contain features closely clustered about a mean size. Thus, determination of the *scale structure*, or size distribution of component features, is important in characterizing random interfaces for acoustical applications. However, because the PSD is only a second moment characterization, a power law PSD can allow both single-scale and multiscale processes. Only if the process is Gaussian is the PSD a complete description. Wavelet representations succeed where Fourier methods fail, as they are ideally suited for determining scale structure. In the case of fractal interfaces, the wavelet coefficients are Gaussian and independent across scale and space. Single-scale interfaces lead to a dependence of wavelet coefficients across scale for a given spatial location and vice-versa.

Session 1pPA

Physical Acoustics: Thermoacoustic Engines

William P. Arnott, Chair
Desert Research Institute, Atmospheric Sciences Center, P.O. Box 60220, Reno, Nevada 89506-0220

Contributed Papers

1:30

1pPA1. Thermoacoustic measurements on a single pore with an applied temperature gradient. Larry A. Wilen, Gabriela Petculescu, and Andi G. Petculescu (Dept. of Phys. and Astron., Ohio Univ., Athens, OH 45701, wilen@helios.phy.ohiou.edu)

Previous measurements [L. Wilen, J. Acoust. Soc. Am. **101**, 3022(A) (1997)] probed the geometry-dependent thermal coupling between a gas and the walls of single pores over a wide range of ratios of the thermal penetration depth to pore size. A volume modulation method was described which allows one to measure the complex compressibility of the pore directly. The present work extends those results to pores with an applied temperature gradient. Unlike the earlier experiments, the results are sensitive to terms in the thermoacoustic equations which depend on a gradient, and the theoretical predictions can be tested in a direct way. After a brief review of the experimental technique, data for pores of different lengths and gradients will be presented and discussed. Possible new applications of the technique will be proposed. [Work supported by Ohio University Research and Sponsored Programs.]

1:45

1pPA2. Experimental investigation on the transition to steady state of self-oscillation of a thermoacoustic prime mover. Bosen Zhao, Fathi Jebali, and Maurice X. Francois (LIMSI, CNRS, B.P. 133, F-91403 Orsay, Cedex, France)

A thermoacoustic prime mover can be considered as a nonlinear system in which an initial small perturbation is amplified until nonlinear mechanisms lead to a new balanced state. A series of experiments were carried out to help understand the nonlinear mechanism. The transition process is divided into three periods: a growing period in which the pressure grows exponentially with time to its maximum; a falling period in which the pressure decreases to its minimum; and a period in which the pressure is restablished slowly until the steady state. A rapid decrease in temperature difference between the cold and hot heat exchangers is observed when the pressure arrives at the maximum. The difference between maximum pressure and steady-state pressure depends on the mean pressure in the prime mover: the higher the mean pressure, the greater the difference. A model of positive feedback amplifier with nonlinear resistance is proposed to describe this phenomenon.

2:00

1pPA3. Improvements in an experimental thermoacoustically driven thermoacoustic refrigerator. Thomas J. Hofler and Jay A. Adeff (Dept. of Phys., Naval Postgrad. School, Code PH/HF, Monterey, CA 93940, hofler@physics.nps.navy.mil)

A thermoacoustically (heat) driven, thermoacoustic refrigerator apparatus, having a novel topology, has produced significant cooling power and efficiency. It has achieved a cooling temperature span of 60 °C, and 91 W of cooling power at a span of 25 °C, with an overall COP of 0.15. These numbers were produced with a porous carbon refrigerator stack and a driver stack comprised of stainless steel wire mesh disks. While easy to fabricate, the wire mesh stack is known to produce poor efficiency with the commercially available wire sizing. New modifications currently being fabricated include: a stainless steel foil roll stack for the driver, a plastic film roll stack for the refrigerator, as well as improved heat exchangers. Substantial increases in amplitude, cooling power, and overall COP are anticipated as a result of these modifications. [Work supported by Office of Naval Research.]

2:15

1pPA4. Numerical optimization of a novel thermoacoustically driven thermoacoustic refrigerator. Thomas J. Hofler (Dept. of Phys., Naval Postgrad. School, Code PH/HF, Monterey, CA 93940)

Previous modeling work on a novel heat-driven refrigerator has generated overall COP numbers in the range of 0.45 using pure helium gas. The modeled conditions are for commercial or air-conditioning cooling temperatures, with a hot drive temperature of 400 °C. The method used for these results involved the manual optimization of a modest number of engine parameters. New modeling work utilizes optimization algorithms in conjunction with a larger parameter set. A more aggressive choice of operating conditions including binary gas mixtures and higher drive temperatures are also explored, resulting in much higher overall COP values. [Work supported by the Office of Naval Research.]

2:30–2:45 Break

2:45

1pPA5. Numerical investigations of a two stack annular prime mover. Hsiao-Tseng Lin (Vehicle Eng. Dept., Chung Cheng Inst. of Technol., Taiwan), Ralph T. Muehleisen (Dept. of Phys., Naval Postgrad. School, Monterey, CA 93943), and Anthony A. Atchley (Penn State Univ., University Park, PA 16802)

A numerical analysis of the two-stack thermoacoustic prime mover in an annular resonator has been made. It was found that the addition of a second stack significantly alters the eigenmodes of a single stack annular prime mover such that thermoacoustic growth may be supported. Simulations predict that one mode of the two-stack prime mover could reach onset at a reasonable temperature. The performance of the two-stack prime mover with various spacings between the stacks is presented. The results are compared to the performance of single-stack annular prime mover in a constricted annular resonator. [Work supported by the Office of Naval Research and the American Society for Engineering Education.]

3:00

1pPA6. Design of a thermoacoustic refrigerator for visualization measurements. Martin Wetzel and Cila Herman (Dept. of Mech. Eng., Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218, herman@titan.me.jhu.edu)

In previous studies, holographic interferometry (HI) combined with high-speed cinematography was applied to investigate the oscillating temperature field in a thermoacoustic refrigerator model [Wetzel *et al.*, J. Acoust. Soc. Am. **100**, 2846(A) (1996)]. While important flow parameters of thermoacoustics could be matched in these experiments, it was not possible to model all length scales because of design requirements imposed by HI. In order to resolve this problem, a novel design algorithm was developed based on the short-stack boundary-layer approximation.

The advantage of this algorithm is that all design parameters are presented in normalized form, and therefore, the design optimization can be performed in the normalized multidimensional parameter space without determining the actual dimensions of the device. Once this optimization has been completed, the actual dimensions can be determined according to the design requirements imposed by HI. At the meeting, the design as well as initial measurements with the thermoacoustic refrigerator will be presented. [Work supported by the Office of Naval Research.]

3:15

1pPA7. Analogy between the circular acoustic waveguide with axial temperature gradient and the electrical transmission line with source and loss. Bosen Zhao, Fathi Jebali, and Maurice X. Francois (LIMSI, CNRS, B.P. 133, F-91403 Orsay, Cedex, France)

A current source introduced by an axial temperature gradient in a circular acoustic waveguide has been examined. It is shown that the effect of the axial temperature gradient is to amplify (or attenuate) the acoustic power in the waveguide where losses due to viscous effect and thermal effect are described, respectively, by a series impedance per unit length and a shunt admittance per unit length. This analogy is very useful for calculating transmissions for cases such as thermoacoustic machines, automotive mufflers, or pulse combustors. Examples of application for a thermoacoustic prime mover calculation and its experimental comparison are presented.

3:30

1pPA8. Determination of source data for a thermoacoustic prime mover by the multiload method. Bosen Zhao, Fathi Jebali, and Maurice X. Francois (LIMSI, CNRS, B.P. 133, F-91403, Orsay, Cedex, France)

A thermoacoustic prime mover can be used as a sound source to drive thermoacoustic refrigerators. The nitrogen-filled prime mover studied is a quarter-wavelength resonator that produces sound at nominally 86 Hz for a temperature difference of $\Delta T \geqslant 430$ K. The pressure and specific impedance at the mouth of the prime mover were measured as a function of the load realized by different reservoirs. These measurements allow one to determine the characteristics of the prime mover at operation states and therefore to predict the maximum acoustic power at a perfect match. Nonlinear losses in the prime mover are analyzed by estimating the measured impedance.

MONDAY AFTERNOON, 1 DECEMBER 1997 CHAMBER ROOM, 2:00 TO 4:30 P.M.

Session 1pSA

Structural Acoustics and Vibration: Acoustic Scattering from Elastic Structures

John J. McCoy, Chair
School of Engineering, The Catholic University of America, Washington, DC 20064

Contributed Papers

2:00

1pSA1. Local admittance fluctuations and the scattering from a cylindrical shell with many internal oscillators. J. A. Bucaro, D. M. Photiadis, B. H. Houston (Naval Res. Lab., Code 7130, 4555 Overlook Ave. S.W., Washington, DC 20375-5350, jbucaro@ccf.nrl.navy.mil), and A. J. Romano (Sachs Freeman Assoc., Landover, MD 20774-5322)

The angular and spectral scattering patterns from local admittance fluctuations, assumed to be caused by the presence of mechanical, dumbbell oscillators attached more or less randomly in the interior of a cylindrical shell, are predicted numerically and compared to the experimental measurements reported by Bucaro *et al.* [Proc. ASME, Noise Control and Acoustics Division, NCA **22**, 87–92 (1996)] and Photiadis *et al.* [J. Acoust. Soc. Am. **101**, 895–899 (1997)]. The numerical results are shown to be in reasonable agreement with what is observed experimentally. It is suggested that scattering contributions from such local admittance fluctuations can be a significant component in the midfrequency scattering patterns of submerged targets of interest. [Work supported by the Office of Naval Research.]

2:15

1pSA2. The radiation and vibratory response of a fluid-loaded structure with high internal complexity. B. H. Houston, D. M. Photiadis, and J. A. Bucaro (Naval Res. Lab., Code 7130, 4555 Overlook Ave. S.W., Washington, DC 20375-5350, houston@lpsa1.nrl.navy.mil)

Recently, experiments were reported in which complex internal structure was shown to dramatically influence the scattering cross section of fluid-loaded shells [Photiadis *et al.*, J. Acoust. Soc. Am. **101**, 895–899 (1997)] and where spatially varying local admittances were proposed to explain the observed scattering details [Bucaro *et al.*, J. Acoust. Soc. Am. **100**, 2721 (1996)]. Here, measurements of the *vibratory* response of this same structure [a ribbed shell with a large number ($\sim$1000) of internal oscillators] are shown. Comparisons are made to the response of an identical shell with no internal oscillators. For the complex structure, no distinct pass and stop bands for any of the circumferential orders is seen, and a high degree of localization at all circumferential orders and frequencies is observed. Generally speaking, the wave-number-frequency plots associated with the complex structure are significantly different than those for the simple framed cylinder at all but the highest azimuthal components ($n > 17$). Moreover, the wave-number decompositions of these data reveal a strong dominance of local bending in the response of the complex structure and evidence that strong coupling between the circumferential orders contributes to the significantly enhanced radiation levels. [Work supported by the Office of Naval Research.]

2:30

1pSA3. A multiresolution analysis of scattering by a pair of local regions of complex heterogeneity. John J. McCoy (Catholic Univ. of America, Washington, DC 20064) and Ben Z. Steinberg (Tel Aviv Univ., Tel Aviv, Israel)

The scattering applies to flexure waves in a beam. Each of the local regions have a spatial extent that is a small multiple, say 2 or 3, of the length of the waves in the homogeneous background beam. The complexity of the heterogeneity refers to an irregular variation on a length scale that is a small fraction, say 1/100, of the size of the scattering region. A multiresolution analysis refers to a two-step solution methodology, by which an ''effective'' property description of the scatterers is first obtained on accomplishing a formulation substructuring in a wavelet-based phase space. The spatially filtered component of the scattered field, i.e., the only component that obtains outside the regions of heterogeneity, is then obtained via a formulation expressed in this effective description. Demonstrated, both analytically and numerically, is that the interaction of

the scattering regions is accommodated in the second step of the solution procedure. Thus the effective description of each of the regions is that which applies in the absence of the second region, provided the regions do not overlap.

2:45

1pSA4. Local modeling of complex elastic structures. Douglas M. Photiadis (Naval Res. Lab., 4555 Overlook Ave. S.W., Washington, DC 20375)

Several recent results, both numerical and experimental, have indicated that the structural acoustic response of large vibratory systems with complex internal structure is dominated by local response in the midfrequency range. These results have clear implications regarding control strategies for noise control, but also may have a significant impact on the acoustic scattering characteristics of complex structures. In this paper, this phenomena is analyzed theoretically using two disparate models, a conventional stochastic wave scattering model and a new type of model, which treats the system instead as a set of coupled, local resonators. This local model more naturally accommodates the observations concerning the generally localized response but is somewhat more difficult to associate with the geometrical characteristics of the ''base structure'' than a more conventional elastic wave scattering model. A number of aspects involving the vibration are analyzed and contrasted between the two models; the occurrence of simple, relatively isotropic spreading of energy, increased damping resulting from the internals, increased damping resulting from radiation into the surrounding fluid, and last but not necessarily least, the impact of local resonance behavior.

3:00

1pSA5. Feature extraction based on eigenvector analysis applied to the monostatic scattered field of a ribbed finite cylinder. Angie Sarkissian (Naval Res. Lab., Washington, DC 20375, angie@aqua.nrl.navy.mil)

Feature extraction based on eigenvector analysis is applied to the monostatic response of a ribbed finite cylindrical shell with hemispherical end caps. The method, which is based on the Karhunen–Loeve expansion, is applied in the frequency domain to extract features for optimal representation of the data. An orthonormal set of eigenvectors that form a set of basis functions are computed by diagonalizing the correlation matrix. The expansion of the monostatic scattered field with the resultant set of basis functions is optimal because a small number of the functions is required to approximate the scattered field at each orientation of the scatterer. Such a representation reduces the dimensionality of the problem by more than an order of magnitude. The method is applied to two frequency ranges. In the first case, enhancements in target strength are present due to the phase matching of the elastic waves to the acoustic waves in the exterior fluid. In the second case, Bloch wave resonances are present due to the periodicity of the ribs. It is shown that as larger variations are present in target strength as a function of frequency, a larger number of eigenvectors are necessary to approximate the scattering response.

3:15–3:30 Break

3:30

1pSA6. Angle dependence of the meridional leaky-ray backscattering enhancement from the end of a tilted finite cylinder. Convolution analysis and a numerical test for shells. Philip L. Marston and Scot F. Morse (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

Reflection of a meridional leaky ray from the far end of a tilted cylinder produces a backscattering enhancement when the tilt angle is close to the leaky wave coupling angle. The ray lies in the meridional plane defined by the incident wave vector and the cylinder's axis. The peak magnitude of this enhancement was related to the end-reflection coefficient by a convolution surface integral [P. L. Marston, J. Acoust. Soc. Am. **102**, 358–369 (1997)]. The more difficult integrals descriptive of the dependence on tilt angle are numerically evaluated in the present study. Though the exact analytical solution of the problem considered is unknown, for comparison an approximate partial-wave series description is available where the boundary conditions at the ends of the cylinder are such that the end-reflection coefficients become unimodular. For antisymmetric leaky Lamb waves on long thick and thin shells, the enhancement width, magnitude, and location each agree with the result from the integral when a unimodular reflection coefficient is also assumed in the convolution formulation. The convolution analysis should also apply to meridional leaky Rayleigh waves reflected by the ends of a solid cylinder. [Work supported by the Office of Naval Research.]

1pSA7. Abstract withdrawn.

3:45

1pSA8. Observations of backscattering of obliquely incident plane waves by composite cylindrical shells constructed from isotropic and transversely isotropic layers. Gregory Kaduchak (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

Acoustic backscatteirng from finite, composite cylindrical shells in water is examined. The shells are comprised of N layers, which may be described by orthotropic or transversely isotropic materials. The present research examines experimental observations of the scattering signatures obtained from obliquely incident plane waves in the mid- and high-frequency regions. Scattering effects are viewed in both the time and frequency domains. Attention will be given to the similarities (and dissimilarities) of the scattering signatures, which are the chief contributors to the backscattering form function as the symmetry axis of the transversely isotropic layers is rotated away from the axial direction. To localize the sources of scattered radiation at oblique incidence, scattering effects are viewed with high-resolution techniques, which include narrow transmit/receive beams as well as synthetic aperture sonar. [Work supported by the Office of Naval Research.]

4:00

1pSA9. Applications of the causality condition to acoustic scattering. J. Gregory McDaniel (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

The causality condition states that the response of a passive system cannot precede the cause. Under certain conditions, the causality condition leads to a Hilbert transform relation between the magnitude and phase of the complex Fourier transform of a system's response. This relation has profound implications for those attempting to design passive structures whose desired scattering characteristics are expressed in the frequency domain. Unless the causality condition is satisfied in the frequency domain, the structure is not physically realizable. In this presentation, some novel applications of this relation are developed for a one-dimensional fluid-loaded structure which scatters incident sound in the backward and forward directions. In each application the reflection and transmission coefficients, which are the complex Fourier transforms of the reflected and transmitted pressures due to an impulsive incident pressure wave, are subject to the causality condition. The Weiner–Lee transform, which is derived from the Hilbert transform but is more easily implemented numerically, is used to find complex reflection and transmission coefficients given only their frequency-dependent magnitudes. By using this informa-

tion and structural reciprocity, one can find an impedance matrix of a structure which scatters sound in a specified way.

4:15

1pSA10. Acoustic excitation of generalized normal-mode vibrations on hemispherically endcapped and on infinite elliptic cylinders. G. Maze, J. Lanfranchi, D. Décultot, J. Ripoche (LAUE, URA CNRS 1373, Univ. of Le Havre, France), and H. Überall (Catholic Univ., Washington, DC 20064)

The concept of normal-mode vibrations of elastic objects, and their resonant acoustic excitation, can be straightforwardly demonstrated analytically for objects of canonical shape (spheres, circular cylinders) only, but it evidently remains valid for objects of completely general shapes. Mode analysis may be extended to these by invoking the principle of phase matching of surface waves (circumferential waves) that encircle the object after their acoustic excitation [H. Überall, L. R. Dragonette, and L. Flax, J. Acoust. Soc. Am. **61**, 711 (1977)], thereby creating standing waves around the object (resonances) that correspond to its normal modes of vibration. Experiments have been performed leading to the excitation of resonances of submerged spherically-endcapped circular cylinders, and of infinite elliptic cylinders, which were interpreted according to the phase-matching concept, thus leading to a description of their vibrations as generalized normal modes.

MONDAY AFTERNOON, 1 DECEMBER 1997 SENATE/COMMITTEE ROOMS, 1:30 TO 5:00 P.M.

Session 1pSC

Speech Communication: Workshop: Basic Science at the Intersection of Speech Science and Communication Disorders

Lynne E. Bernstein, Cochair
Spoken Language Processing Laboratory, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

Gary Weismer, Cochair
Waisman Center, University of Wisconsin, 1500 Highland Avenue, Madison, Wisconsin 53705

Continuation of Workshop—See Session 1aSC page 3068

Session 1pSP

Signal Processing in Acoustics and Structural Acoustics and Vibration: Signal Processing for Multi-Channel Vibrational Analysis

David J. Evans, Chair
National Institute of Standards and Technology, Building 233, Room A147, Gaithersburg, Maryland 20899

Chair's Introduction—1:00

Invited Papers

1:05

1pSP1. New techniques for nonlinear system analysis and identification from random data. Julius S. Bendat (J. S. Bendat Co., 833 Moraga Dr., No. 10, Los Angeles, CA 90049)

Direct and reverse MI/SO (multiple-input/single-output) techniques are new frequency-domain techniques that provide accurate practical methods to analyze and identify the dynamic properties of nonlinear systems. The Dirac MI/SO technique is applicable to nonlinear systems with specified parallel linear and nonlinear transformations. The reverse MI/SO technique is applicable to nonlinear systems that can be reasonably modeled by nonlinear integrodifferential equations of motion. Nonlinear systems are included where the coefficients can be constants or frequency dependent. Simulated or measured data can have arbitrary probability and spectral features. Each of the identified nonlinear components can be evaluated at any desired frequency with separate coherence functions. Thus these techniques represent a significant advance in using real-world data to help improve the design and understanding of nonlinear systems. This presentation will review the analytical basis of these new techniques and illustrate their application from material in the latest book by J. S. Bendat [*Nonlinear System Techniques and Applications* (Wiley-Interscience, New York, 1998)].

1:45

1pSP2. Measurement and analysis of structural wave types on fluid-loaded shells. David Feit and David C. Warwick (Carderock Div. Naval Surface Warfare Ctr., 9500 MacArthur Blvd., West Bethesda, MD 20817)

The frequency-wave-number spectrum of the vibration field of force excited, fluid-loaded cylindrical shells is a significant determinant of the shells' acoustic radiated field. This presentation discusses the procedures used and results obtained using accelerometer arrays to sample the vibrations on a number of different type shells. The arrays are formed by distributing accelerometers with uniform spacing, either along a longitudinal generator or around a circumference of the cylinder. Depending on the orientation of the accelerometers relative to the shell's midsurface, the dispersion characteristics found allow for the identification of the various wave types that exist on fluid-loaded cylindrical shells. These are the quasi-flexural waves and in-plane membrane waves, both longitudinal and shear. In an early set of measurements, made in 1986, the shell structure could be considered as unstiffened, while the other results show the effects of stiffening ribs in the frequency-wave-number plots. These effects manifest themselves as aliased quasiflexural waves. The shapes of the dispersion curves are in good agreement with numerical results arising from analytical expressions.

2:15

1pSP3. Power flow in structures. Gunnar Rasmussen (G.R.A.S. Sound & Vibration aps., Skelstedet 10B, 2950 Naerum, Denmark)

Power released in or transferred to a structure will cause noise radiation, fatigue, or mechanical malfunction in the structure or parts connected to the structure. Nonlinear behavior is of great significance to the integrity of any structure and is therefore of fundamental interest. An ideal machine would produce no vibration at all because all energy would be channeled into the job to be done. In practice, vibration occurs as a by-product of the normal transmission of cyclic forces through the mechanism. Machine elements react against each other and energy is dissipated through the structure in the form of vibration and acoustic noise. If the surface area is large compared to an acoustic wavelength, acoustic intensity is a good indicator of dynamic activity. If the area is small compared to a wavelength at the frequencies in question, surface measurements using the vibratory motion on the structure's surface will be a good indicator. If the energy transfer takes place between two separate structures coupled at points, the point power measurement method should be applied. Two accelerometers, accelerometer and force gauge, two microphones, multi microphone arrays or laser techniques may be used for transduction. Signal processing and interpretation will be discussed.

2:45

1pSP4. Order tracking with multiple shafts and crossing orders. Håvard Vold (Vold Solutions, Inc., 1716 Madison Rd., Cincinnati, OH 45206)

Transient sinusoids that cross in frequency exhibit interaction phenomena that defy conventional tracking filter techniques. The Vold–Kalman filter explicitly uses shaft speed information from multiple shafts to decouple interacting orders through a simultaneous estimation. The estimation uses energy constraints to distribute the total signal energy between the orders. The outputs of the filter are time histories of complex envelopes of the order functions, without beating interactions and with no phase bias. The estimation is also independent of slew rates, such that highly transient events may be tracked with high fidelity. By modulating the complex envelopes

by suitable carrier waves, time histories of the individual orders are obtained which can be used for sound quality studies and sound synthesis. The phase information allows for the construction of operating deflection shapes as a function of shaft, order, and shaft speed. Examples will be shown for vibrations in continuous rate belt drive transmissions and sound synthesis for torque converters.

3:15–3:30 Break

Contributed Papers

3:30

1pSP5. A comparison of partial coherence and singular value partial field decomposition in the context of near-field acoustical holography. Hyu-Sang Kwon, J. Stuart Bolton (1077 Ray W. Herrick Labs., School of Mech. Eng., Purdue Univ., West Lafayette, IN 47907-1077), and J. K. Hammond (Univ. of Southampton, Southampton SO17 1BJ, England)

Sound fields radiated by complex noise sources, e.g., automotive engines, generally comprise superposed individual fields generated by incoherent subsources. When applying near-field acoustical holography (NAH) to such fields it is first necessary to decompose the total sound field measured on the hologram plane into coherent partial fields, each of which is then projected to a reconstruction plane where they are summed on an energy basis. The partial field decomposition is performed after first calculating the cross spectra between a number of reference microphone signals that have been conditioned by using either partial coherence or singular value decomposition procedures, and the spatially sampled sound pressures on the hologram plane. That cross-spectral information then allows the partial fields to be created. In this presentation, the performance of the two reference signal decomposition procedures will be compared theoretically and experimentally. In particular, it will be shown that partial coherence decomposition allows the partial fields to be associated with the fields radiated by individual subsources if the latter are sufficiently separated spatially. In contrast, the singular value decomposition procedure usually results in partial fields that combine the properties of the subsources, and are thus not directly related to the fields radiated by individual subsources.

3:45

1pSP6. Vibroacoustical multichannel diagnostics of energetic systems. Leonid M. Gelman (Dept. of Nondestructive Testing, Natl. Tech. Univ. of Ukraine, 37, Peremogy pr., Kiev, 252056, Ukraine)

A new multichannel low-frequency vibroacoustical diagnostics method of energetic systems is considered theoretically and experimentally. The proposed method represents one of the approaches to efficient energy use and conservation. This method is based on processing of the low-frequency vibroacoustic noise of energetic systems. The main advantages of the method: early preventive nondemountable automatic diagnostics on all operating regimes of energetic systems.

4:00

1pSP7. Spectral function of the shocks of acoustic fluctuation phenomena. Dat H. Tran and Alexsandr J. Krasilnikov (Dept. of Electron., Ukrainian Natl. Tech. Univ., Kiev, 252057, Ukraine)

In the description of the acoustic fluctuation phenomena (AFP), for example, sonocavitation noise, the signals of the acoustic emission, scattering on the heterogeneity, reverberation, and others, the model of shot noise is often applied and the principle of distribution is considered to be Gaussian. This model of AFP is proposed and proven with the conjecture that the shape of the impulse depends on the time of its generation, and amplitude of the impulse is subjected to the inhomogeneous Poisson process. It was shown that the characteristic function of the linear process is an unlimited dividend and, therefore, the function of the distribution or the probability density cannot be obtained in the evident mode, even through the mixed function. The spectral function of shocks (SFS) was proposed for the investigation of the principle of distribution and for the determination of the characteristic function. The algorithm was obtained that permits finding the SFS. Komologov, Levin, and some properties of SFS as well and the results of the calculation of SFS of the different AFPs are shown.

4:15

1pSP8. Statistical characteristics of the acoustic emission signals. Dat H. Tran and Alexsandr J. Krasilnikov (Dept. of Electron., Ukrainian Natl. Tech. Univ., Kiev, 252057, Ukraine)

The typical block of signal processing in the instrument of diagnostic and construction damage prediction by the acoustic emission (AE) method was proposed, which consists of an inertial detector, comparator, and counter. The correlation between the registered parameters (number of impulses, activity, average level, density of amplitude distribution) and the initial AE flux on the input of the signal was described by the method of linear random processes. On the basis of phenomenon modeling, it was shown that, in an ideal condition, the discrete AE signal fits the gamma distribution. The dynamical approach to the model also attempts to extend it in relation to the three-dimensional coordinates. The deviations of the relation between the registered and proper AE meanings and the method of the results correction are shown.

4:30

1pSP9. Vibroacoustical multichannel nondestructive evaluation method of fatigue cracks. Nadezhda I. Bouraou (Dept. of Orientation and Navigation, Natl. Tech. Univ. of Ukraine, 37, Peremogy pr., Kiev, 252056, Ukraine), Leonid M. Gelman, and Natalia Yu. Ossokina (Natl. Tech. Univ. of Ukraine, Kiev, 252056, Ukraine)

For nondestructive testing (NDT) and nondestructive evaluation (NDE), the low-frequency vibroacoustical free-oscillation method is used. For the first time, a theoretical investigation is carried out to differentiate between spectral density and decrement of free oscillations of testing objects in the presence and the absence of fatigue cracks taking into account the internal friction in the object material. The new analytical dependencies of the spectral density and decrement of testing object-free oscillations from the relative crack size, factor of internal friction, duration of NDT(E), and initial object speed are received, on the basis of which the new multidimensional vector of testing data for NDT(E) is proposed. Experimental results with aircraft engine blades in the presence and the absence of fatigue cracks, which match with theoretical results, are presented. The received results are common; therefore, they have been found to be expedient in taking into account the obtained results for forced oscillation NDT(E) methods.

Session 1eID

Interdisciplinary: Tutorial Lecture on The Dolphin Echolocation System

Alexandra I. Tolstoy, Chair
Integrated Performance Decisions, 224 Walalae Avenue, Suite 5-260, Honolulu, Hawaii 96816

Chair's Introduction—7:00

7:05

1eID1. The dolphin echolocation system. Whitlow W. L. Au (Marine Mammal Res. Program, Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734)

The sonar of dolphins may be the most sophisticated of all sonar systems, biological or man-made, in shallow waters and for short ranges. The Atlantic bottlenose dolphin emit short-duration (50–70 μs), high-frequency (120–140 kHz), broadband (40–50 kHz) echolocation signals with peak-to-peak amplitudes up to 228 dB re: 1 μPa. The type of signals used by dolphins play a significant role in their sonar discrimination capabilities. They have been observed detecting, classifying, and retrieving prey that is buried in sandy bottom up to a depth of about 0.3 m. In addition, controlled echolocation experiments have shown that dolphins can discriminate wall thickness, material composition, shape, and size of targets. The echolocation system of dolphins will be discussed in three parts. The first part will consider the properties of the receiving system (auditory capabilities). The second part will deal with the characteristics of echolocation signals, and the third part will consider several sonar capabilities.

Session 2aAA

Architectural Acoustics and Engineering Acoustics: Loudspeakers for Listening Spaces

Neil A. Shaw, Chair
Menlo Scientific Acoustics, Inc., P.O. Box 1610, Topanga, California 90290-1610

Chair's Introduction—8:30

Invited Papers

8:35

2aAA1. The control of beam direction and beamwidth of loudspeaker arrays. Jefferson A. Harrell (Jet Propulsion Lab., Pasadena, CA) and Elmer L. Hixson (ECE Dept., Univ. of Texas, Austin, TX 78712)

By the control of time delay and the frequency function of each speaker in an array, the direction and coverage as a function of frequency can be controlled. Beamwidth can be held constant with frequency so that everyone in the coverage area will receive the same spectrum of sound. Examples of one-dimensional beam control will be presented using broadside and endfire line arrays. Then two-dimensional coverage control with crossed broadside and stacked endfire arrays will be demonstrated. Advantages of controlled loudspeaker arrays will be discussed.

9:00

2aAA2. Loudspeaker array designs using drivers positioned to improve directionality. Christopher ''Topper'' Sowden (Pelton Marsh Kinsella, 1420 Elmbrook, Ste. 400, Dallas, TX 75247)

This paper provides a study of the design methodology used in the development of a central loudspeaker cluster for stadiums and arenas. The design criteria is compared with the actual constructed results to describe the effective use of the concept and use in future designs. The methodology for the design is based on the paper published in the Journal of the Audio Engineering Society in April of 1990 by Jefferson Harrell and Elmer Hixson. The paper by Harrell and Hixson discussed horizontal and vertical directionality of multidriver loudspeaker arrays when spaced a calculated distance from each other. This paper provides a demonstration of these concepts with systems recently completed at the Bryce Jordan Center at Penn State University, and Sun Devil Stadium at Arizona State University. The methodology was applied to an array with transducer spacings greater than 2 m and improved control of frequencies down to 50 Hz. Digital control of audio provides loudspeaker array control and alignment capabilities not previously possible through analog techniques. The paper discusses adjustment of the arrays installed at the above-mentioned facilities.

9:25

2aAA3. An improved constant directivity loudspeaker array. Vance Breshears (Sound Technol. Consultants, 2707 Via Viejas, Alpine, CA 91901), Ralph Heinz, and Brian Oppegaard (Renkus-Heinz, Irvine, CA 92614)

A simple design goal for any loudspeaker cluster would be to provide even sound pressure levels and consistent frequency response for any seat in a listening space. To achieve this goal, many clusters are currently being designed around trapezoidal cabinets promising array ability. However, the resulting lobing, beaming, and comb filtering these types of arrays typically produce, make it nearly impossible to realize the design goal stated above. Even the best implementations of this type, using a trapezoidal cabinet or horn with consistent dispersion characteristics, result in a system whose low-frequency dispersion in the vertical plane does not match the level of controlled directivity of the mid- and high-frequency range. A new method has been developed that provides a systems approach to building coherent point source clusters with frequency invariant coverage from below 100 Hz to 18 kHz in both horizontal and vertical planes.

9:50–10:00 Break

10:00

2aAA4. High-performance loudspeaker driver design for indoor/outdoor applications. William J. Kieltyka (New England Audio Resource Corp., 12 Foss Rd., Lewiston, ME 04240, near@maine.com)

The design and manufacture of drivers to provide the utmost in sound quality and intelligibility and to offer weatherproof qualities is the challenge. Conventional dynamic drivers with treatment can offer limited protection, but the sound reproduction suffers markedly. An approach is used whereby each part can be chosen for performance and reliability when subjected to the worst elements. Rubber compound surrounds, cast frames, metal–alloy cones, high-tech ferrofluids, elimination of centering spiders, and magnets with special plating, are all considerations for improved design. Centering spiders are one critical component where change in mechanical characteristics is bound to occur with time. Subjected to the elements of heat, cold, humidity, etc., aging is accelerated. A speaker cannot perform for any reasonable length of time under these conditions. A design can be used that does not require the spider. When

it is eliminated, the voice coil guidance system (VCGS) becomes the biggest challenge. How to make the ''spiderless'' driver perform under all traditional uses in commercial or consumer sound applications becomes another problem. The proper use of ferrofluid and specially formulated surrounds assures consistent alignment and linear movement. The additional benefits: elimination of spectral contamination caused by the spider; possible improvement in production times.

10:25

2aAA5. Magnetic circuit design methodologies for dual coil transducers used in public sound reinforcement. Douglas J. Button (JBL Professional, 8500 Balboa Blvd., Northridge, CA 91329)

Although the electrodynamic drive topology incorporating two axially (not concentrically) placed coils is not a new idea and dates back to the 1950's, recent interest in the design has begun a proliferation of the design concept in professional sound reinforcement loudspeakers. The fundamental design involves two coils that are opposite in phase and reside in oppositely polarized magnetic gaps, thus providing a Lorentz force in the same axial direction. The two gaps are formed from the same magnetic circuit, and both coils are wound on the same form, separated by some distance, connected or wound out of phase, and attached to a single diaphragm. The work herewithin describes different design options in magnetic materials, magnetic circuit geometries, and voice coil topologies. The focus is on performance trade-offs and advantages in weight, power handling, power compression, and distortion, relative to a single gap design. The end result being acoustical solutions with better fidelity, less weight, and size.

10:50

2aAA6. The use of the personal computer to assist in the design of loudspeaker systems. Robert C. Coffeen (Univ. of Kansas, Architectural Eng., Marvin Hall, Univ. of Kansas, Lawrence, KS 66045) and Gabriel Caunt (Univ. of Kansas, Lawrence, KS 66045)

Through the use of building acoustic analysis and design computer programs, which include the construction of building models within the program, which provide for the installation of loudspeakers within the model, and which include performance data for the loudspeakers being used, the loudspeaker system designer is provided with information and assistance which may not otherwise be readily available. The design of a typical loudspeaker system will be discussed with the aid of the computer program EASE. Then, if an auralization program is available to the designer, the performance of the loudspeaker system may be auditioned with the particular acoustical characteristics of the room in which it is installed as simulated by the building model. The performance of a particular loudspeaker system will be demonstrated using the computer program EARS, with changes to room acoustical characteristics and ambient noise conditions.

Contributed Paper

11:15

2aAA7. Finite-amplitude wave propagation in big sound systems. Gene Czerwinski, Alex Voishvillo, Sergei Alexandrov, and Alex Terekhov (Cerwin Vega, Inc., 555 East Easy St., Simi Valley, CA 93065)

Propagation of large-amplitude sound waves is accompanied by the generation of air nonlinear distortion. This effect as related to audio signals was researched in the thirties by Thuras, Jenkins, and O'Neil. They investigated the waveform distortion of plane waves. The sound waves radiated by loudspeakers are all but plane waves. In real loudspeakers, the ''degradation'' of waveforms depends on such factors as directivity and the placement of loudspeakers. Therefore, the results derived from plane-wave tube propagation are hardly applicable. Since nonlinear distortion of loudspeakers interacts with waveform distortion, it is difficult to distinguish one from the other. Two groups of loudspeakers were used and single tones were put on each group to exclude distortion, inherent to loudspeakers. Therefore, intermodulation products indicative of the air distortion were measured. Intermodulation distortion was measured at different frequencies and SPL with progressively increasing path lengths, and various distances between loudspeakers. It was shown that this distortion is inherent to big sound systems, and the level of distortion can be significant. Waveform distribution was also simulated on computer by solving the KZK equation. Air distortion depends on SPL, frequency, distance, loudspeaker directivity, and placement. Some aspects to minimize air distortion in sound systems are discussed.

Session 2aAO

Acoustical Oceanography and Underwater Acoustics: Stochastic Inverse Methods Applied to Ocean Processes

Terry E. Ewart, Chair
Applied Physics Laboratory, University of Washington, MS HN-10, Seattle, Washington 98105

Chair's Introduction—8:25

Special Lecture

8:30

2aAO1. Inverse problems in radar measurements of stochastic processes in the ionosphere. John Sahr (Dept. of Elec. Eng., Univ. of Washington, P.O. Box 352500, Seattle, WA 98105)

This paper offers a short introduction to certain radar studies of the ionosphere. Emphasis will be placed on applications of ''incoherent scatter'' radars relying upon stochastic Thomson scattering to make remarkably detailed measurements of the temperature, density, composition, and other physical parameters in the Earth's ionosphere. The total backscattered power is proportional to the density, but all the other parameters are derived by detailed inspection of the correlation function of density fluctuations. Measuring the correlation function is quite difficult because of the very small scattering cross section. Furthermore, the ionosphere is ''overspread,'' which means that density fluctuation decorrelates in significantly less time than it takes a radio pulse to transit the region of interest. Simply stated, this means that one cannot measure the correlation function of the ionosphere with a periodic pulse train. Following a very quick tour of the plasma physics and theoretical correlation function (the forward problem) the intricate interplay of the transmitter waveform design and the inverse problem for parameter estimation—including some unique robustness issues—is shown. It will also be shown that the highest performance transmitter waveforms today are themselves stochastic (or nearly stochastic) processes. The results have a pleasing analogy to the quantum mechanics notion of ''observables,'' and shed light on the very nature of measurement.

Invited Papers

9:00

2aAO2. The internal wave spectrum: Evolution, resolution, and inversion. Robert Pinkel (Scripps Inst. of Oceanogr., Univ. of California San Diego, 9500 Gilman Dr., La Jolla, CA 92093)

Internal waves scatter low-frequency sound, complicating the interpretation of long-range acoustic transmission experiments. One can now ''invert'' acoustic data and infer parameters of a modeled internal wave spectrum (S. Flatté, this session). The quality of the result depends, among other things, on the realism of the model spectrum. For many phenomena (surface gravity waves, for example), changes in wave-field energy result in changes in spectral form, as well as level. It is appropriate to investigate this issue with regard to internal waves. Estimates of the vertical wave-number-frequency spectrum of internal waves are presented from six sites around the world (83° N to 2° S). Coded pulse Doppler sonar systems provide the required data. Wave-field energy varies by a factor of 100 in the global sample. The wave-number dependence of the spectrum varies with both wave frequency and overall spectral energy level. These spectral estimates are ''resolved'' in the sense that ocean velocity profiles are inverted to provide estimates of variance density on a band-by-band basis in the wave-number-frequency domain. The ''model mismatch'' issue is not escaped, however, since the waves are often nonsinusoidal. The spectrum alone provides an incomplete description of the field.

9:20

2aAO3. Megameter-scale oceanography of internal waves: Past, present, and future. Stanley M. Flatté (Phys. Dept., Univ. of California, Santa Cruz, CA 95064, smf@pacific.ucsc.edu)

Measurements of internal-wave strengths are reviewed for their contribution to our understanding of megameter-scale behavior. Past measurements are sparse. Recent acoustic measurements have provided large-scale averages, with both temporal and spatial information. Future experiments could provide some insight into the place of internal waves in oceanography.

9:40

2aAO4. Determining the parameters of an ocean internal wave model using acoustic log-amplitude and phase—A Rytov inverse. Steve Reynolds, Terry E. Ewart, and Daniel Rouseff (Appl. Phys. Lab. and School of Oceanography, Univ. of Washington, Seattle, WA 98105)

The feasibility of an inversion for the properties of the stochastic internal wave field using acoustic measurements is demonstrated. Certain parameters of a generalized internal wave model can be obtained for both deep- and shallow-water cases. A hierarchy of inversion scenarios has been studied using numerical simulations where the scenarios vary from extensive vertical and temporal separations to single point measurements. Because acoustic measurements intrinsically average over range, they provide an important

constraint to three-space plus time stochastic models. Using acoustic data with standard oceanographic measurements provides even a more stringent set of inversion constraints. These issues are considered and how acoustics may be used to answer open questions about the properties of internal waves and other random ocean processes is examined.

Contributed Papers

10:00

2aAO5. Just how stochastic are coastal oceanographic fields? John Colosi, Chris Rehmann, and Jim Lynch (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

A key question one must answer before doing an inverse procedure is: ''How deterministic versus stochastic are the fields that are wished to be inverted?'' In this talk, this question is addressed for the coastal processes of solibores, fronts, and coastal eddies, using data from both the SWARM and the shelfbreak front PRIMER experiments, as well as computer models of the processes and acoustic propagation through them. Acoustic propagation over 50-km ranges and frequencies of 50–500 Hz will be considered. Issues concerning the eventual inversion of such data will be discussed in detail.

10:15

2aAO6. New calculations of the acoustic weighting function for ocean internal-wave tomography. John A. Colosi (Woods Hole Oceanogr. Inst., MS 11, Woods Hole, MA 02543)

In acoustic tomography observables like travel time can be expressed in terms of integrals along geometrical optics ray paths where the integrand is called the weighting function. In internal-wave tomography, the travel-time variance is an important acoustic observable and in this case a well-known weighting function has been derived which has the form $F(z_{\rm ray},\theta_{\rm ray})=\langle\mu^2(z)\rangle L_p(\theta,z)/c(z)^2$ where $\langle\mu^2(z)\rangle$ is the fractional-sound-speed variance, $L_p(\theta,z)$ is the effective internal-wave correlation length along a ray-path tangent with angle θ, and $c(z)$ is sound speed. Direct numerical solution for the travel-time variance weighting function shows that acoustic signals are most sensitive to internal waves near the ray upper apex but not as strongly as suggested by F. Examples comparing the direct calculation of the weighting function with F, suggest that the greatest sensitivity to internal waves exists within hundreds rather than tens of meters vertically of the ray upper apex. The inaccuracy of the weighting function, F, can be attributed to the assumptions of small ray curvature and the depth dependence of $\langle\mu^2(z)\rangle$. Comparisons between observations and predictions of travel-time variance for basin-scale acoustic transmissions based on the new weighting function show excellent agreement for internal waves at half the Garrett–Munk spectral level.

10:30

2aAO7. Long-range tomography experiment in the western equatorial Pacific. Gang Yuan, Hidetoshi Fujimori, Toshiaki Nakamura, and Iwao Nakan (Japan Marine Sci. and Technol. Ctr., 2-15 Natsushima-Cho, Yokosuka, 237 Japan)

An ocean acoustic tomography experiment was performed with a pair of 200-Hz transceivers in the western equatorial Pacific along the longitudinal line of 147.5 °E during April–December 1996. Time series of reciprocal ray travel times were obtained at about the 700-km range for approximately 100 days between the transceivers. The one-direction, resolved-ray travel times were applied to obtain temperature, and differential resolved-ray travel times to obtain barotropic current velocity. The inversions derived from the stochastic inverse method gave a series of snapshots of two-dimensional temperature fields, which showed a pattern similar to that determined from CTD measurements during the transceiver deployment cruise. The range-averaged temperature anomaly, $\delta T(z)$, agreed almost perfectly with that calculated from the difference between the CTD temperature profile and the NODC reference temperature profile within an estimated error of ± 0.5 °C, except for the surface layer. The evolution of acoustically determined range-averaged temperature and temperature anomaly structures agreed well with the concurrent PMEL/TAO array data. A range-averaged barotropic current determined from the difference of the reciprocal ray travel times also agreed with the range- and depth-averaged meridional velocity determined from ADCP measurements during the transceiver deployment cruise in the same order.

10:45

2aAO8. Ray/mode duality and turning point filters. Matthew Dzieciuch (Scripps Inst. of Oceanogr., Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093)

In ocean acoustic tomography, the travel time of ray arrivals are used to infer the sound-speed field. In typically dispersive channels, the ray arrivals are followed by energy, which is more easily described by vertical modes. Using a vertical array, the analyst can separate the modal energy in order to construct a dispersion diagram; a measurement of mode group velocity versus frequency for each mode number. The difference between the measured and computed group velocities would then be used in a stochastic inverse for the sound-speed field. In the WKBJ approximation, the expression for mode group velocity can be written as a function of a single parameter, mode axial inclination angle instead of the two parameters, mode number, and frequency. This fact suggests that information from modes with the same inclination angle (or equivalently the same turning point depths), ought to be combined. A turning point filter can be built that does exactly this. Perhaps it is not surprising that nature constructs rays from modes with the same turning point depths also. There is some advantage in this method over standard linear beamforming. [Work supported by the Strategic Environmental Research and Development Program through ARPA.]

11:00

2aAO9. Numerical simulation of the parametric array application for the Fram Strait acoustic monitoring. Konstantin A Naugolnykh, E. C. Shang, and Yun-yu Wang (Univ. of Colorado, CIRES/NOAA/Environ. Technol. Lab, 325 Broadway, Boulder, CO 80303)

An important part of the general problem of the Arctic climate change is the monitoring of the Fram Strait, which lies between Greenland and Spitsbergen. For this reason, it is reasonable to apply acoustic methods which are proven to be effective in acoustic ocean thermometry, acoustic tomography, etc. Knowledge of the main peculiarities of sound signal propagation in this region is needed to estimate the feasibility of applying remote acoustic methods. Due to the extremely complex propagation environment, strong mode coupling takes place which leads to mode repopulation (variation of the modal structure of a sound field with distance). The application of a parametric array as a mode selective source could be useful in this case providing better spatial resolution of the probing area. The results of numerical simulation of parametric array performance in the Fram Strait environment is considered in the present paper.

11:15–11:25 Break

11:25–12:00

Panel Discussion

Session 2aEA

Engineering Acoustics: Advanced Techniques in Ultrasonic Non-Destructive Evaluation

Robert D. Finch, Chair
Cullen College of Engineering, Department of Mechanical Engineering, University of Houston, Houston, Texas 77204-4792

Chair's Introduction—8:25

Invited Papers

8:30

2aEA1. Acousto-ultrasonic nondestructive evaluation of materials and structures. Henrique Reis (Dept. of General Eng., Univ. of Illinois, 117 Transportation Bldg., 104 S. Mathews, Urbana, IL 61801)

A review of the nature of the underlying rationale of the acousto-ultrasonic methodology is presented. The term acousto-ultrasonics denotes a nondestructive evaluation technique that combines some aspects of acoustic emission methodology with ultrasonic simulation of stress waves. Unlike most nondestructive evaluation techniques, acousto-ultrasonics is less concerned with flaw detection than with the assessment of the collective effects of the various flaws and materials anomalies. Acousto-ultrasonics is primarily concerned with material properties variations, such as the significant reduction in strength and toughness caused by combinations of minor flaw and diffuse flaw populations. Porosity content, and fatigue and impact damage are typical examples of factors that affect material properties variations. Applications of the acousto-ultrasonic approach to the nondestructive testing and evaluation of steel-belted radial truck tires and to other structural systems are presented and discussed.

8:55

2aEA2. Acousto-ultrasonic technique for nondestructive evaluation of composites and structures. P. K. Raju (Dept. of Mech. Eng., Auburn Univ., Auburn, AL 36849)

Acousto-ultrasonics (AU) is a nondestructive testing technique for evaluating materials and structures. The name is based on the fact that principles of acoustic emission and ultrasonics are combined in this technique. Unlike conventional ultrasonics, acousto-ultrasonics is not concerned with the detection and characterization of individual flaws. Instead, it is concerned with evaluation of integrated effects of diffuse subcritical flaw populations and microstructural aberrations in composites. Porosity, fiber breaks, resin richness, poor curing, and fiber/matrix bonding are some examples of these, which in turn govern mechanical properties and durability of composite structures. These factors also influence AU signals that consequently reveal that variations in mechanical properties such as tensile strength, stiffness, toughness, etc. [A. Vary, ''Acousto-Ultrasonics,'' *NDT of Fiber Reinforced Plastics Composites*, Vol. 2 (Elsevier Applied Science, 1990). This paper reviews the current state of the art of this technique. Some examples on the use of AU for characterization of composites and evaluation of the adhesive bond strength of composite structures will be discussed. Finally, some of the advantages and limitations of this technique will also be discussed.

9:20

2aEA3. Application of ultrasonic stress measurement to engineering components. Don E. Bray (Dept. of Mech. Eng., Texas A&M Univ., College Station, TX 77843-3123, dbray@mengr.tamu.edu)

Ultrasonic techniques offer several unique advantages in nondestructive stress measurement. Most importantly, the ultrasonic energy propagates below the material surface, giving information on stress distributions away from the surface effects. Additionally, surface preparation requirements are minimal, enabling rapid collection of stress information from several locations in a component. Finally, the instrumentation necessary to collect and analyze the ultrasonic traveltimes required for stress evaluation is available in relatively light, compact, computer driven units that are easily moved into the industrial or laboratory work space. A brief overview will be given of the acoustoelastic effect, which is the basis for ultrasonic stress measurement. Both the acousticbirefringence (shear wave) and the critically refracted longitudinal wave (L_{CR}) methods will be described. Examples of applications of the latter method for stress measurement in turbine components, rolled plate, ductile iron, and railroad rail and wheels will be described.

9:45

2aEA4. A local algorithm for the determination of interior material parameters from a knowledge of elastic displacements. A. J. Romano, J. J. Shirron (Sachs Freeman Assoc., 1401 McCormick Dr., Landover, MD 20785, romano@acoustics.nrl.navy.mil), and J. A. Bucaro (Naval Res. Lab., Washington, DC 20375-5350)

A new method is presented for the local determination of material parameters of an isotropic, inhomogenous medium subject to time harmonic vibration. Given a knowledge of the displacements throughout the medium, a novel implementation of a variational formulation is used to determine the ratios λ/ρ and μ/ρ. A theoretical formulation is presented and validated using numerical data obtained from a finite-element method. The results indicate that the method may be applied locally within an inhomogenous medium, and that the corresponding material parameters can be recovered to a high degree of accuracy. Because the method assumes knowledge of the displacements throughout the medium, various approaches to obtaining this information will also be discussed. [This work supported by the Office of Naval Research.]

10:00

2aEA5. Modulation of ultrasound by vibration and its application for flaw detection. Alexander Sutin (Stevens Inst. of Technol., Hoboken, NJ 07030 and Inst. of Appl. Phys., Russia, asutin@stevens-tech.edu) and Dimitri M. Donskoy (Stevens Inst. of Technol., Hoboken, NJ 07030)

The paper gives an overview of a recent advancement in the development and implementation of an innovative nonlinear modulation technique for nondestructive testing. The theoretical and experimental investigations discovered that a weakly or incompletely bonded interface or contact-type defect causes the modulation of a probing high-frequency ultrasonic wave by low-frequency vibration. The vibration varies the contact area modulating the phase and amplitude of a higher-frequency probing wave passing through the interface. In the frequency domain the result of this modulation manifests itself as sideband spectral components with respect to the frequency of the probing wave. Such a modulation effect was observed experimentally for various materials (metals, composites, concrete, sandstones, glass) with various types of defects such as cracks, debondings, delaminations, and microstructural changes. Two modifications of the technique have been developed: vibromodulation (VM) and impact modulation (IM), employing cw and impact-induced vibrations, respectively. The examples of applications of these methods include crack detection in Boeing 767 steel fuse pins, combustion engine cylinder heads, sandstone, and glass. These methods also proved their effectiveness in evaluating bonding conditions in titanium and composite plates, as well as corrosion in reinforced concrete.

10:15–10:30 Break

10:30

2aEA6. Optimization of time delay laws for focusing with phased arrays. Didier Cassereau (Lab. Ondes et Acoust., Univ. Paris VII, Paris, France)

Recently, phased arrays technology has been greatly improved and represents an efficient tool to focus an acoustic beam in various directions, just changing a time-delay law. Due to the extended dimension of the elements of the array, one major difficulty consists in the optimization of the desired time-delay law. In this paper, PASS (phased array simulation software, developed in the laboratory) is used to illustrate how classical algorithms (average, minimum value, center value) can significantly alter the resulting focal pattern. Then an optimization method based on time reversal is proposed. First, the focal point is replaced by a source and the array measures the resulting field. Then, each channel emits the time reversal of the received signals, thus generating an efficiently focused beam. Finally, a cross-correlation method allows to extract from these signals a time-delay law that is proved to be very efficient and adapted to the expected focal point. Various simulations will be presented to focus the beam through a plane interface separating a fluid and a solid, using two different arrays. It will be shown how classical algorithms fail, while the proposed method based on a time-reversal simulation yields very efficiently focused beams.

10:45

2aEA7. Acoustical imaging by time reversal. Didier Cassereau and Mathias Fink (Lab. Ondes et Acoust., Univ. Paris VII, Paris, France)

For the last few years, time-reversal techniques have been proved to be efficient for defect detection in materials without *a priori* knowledge on the defect. In the case of an extended target, an iterative time-reversal experiment allows focusing on the most reflective point of the defect. At this time, it was not possible to obtain an image of the defect after detection. Using PASS (phased array simulation software, developed in the laboratory), the time-reversal sequence is first iterated in order to be able to focus through a cylindrical interface separating a fluid and a solid. Then, the series of temporal signals is used in conjunction with time-delay laws, thus yielding spatial micro-angulations of the focused beam. Finally, the software is used to compute the back-scattered field due to several extended defects. After calibrating the focal spot displacement with the considered microangulation time delay law, it is shown how this technique allows one to obtain an image of the different defects in the material, even though the standard time-reversal method is only able to focus on the most reflective one. Starting from these numerical results, an experimental validation will start soon.

11:00

2aEA8. Lamb waves generation using embedded piezoelectric elements in composite plates. Jamal Assaad, Emmanuel Moulin, and Christophe Delebarre (Dept. of OAE, IEMN, Valenciennes Univ., BP 311 Le Mont-Houy, 59304 Valenciennes, France)

With the aim of designing nondestructive evaluation and health monitoring systems for composite parts of complex structures (e.g., aircrafts), Lamb waves generated using embedded miniaturized transducers seem to be a very promising technique. Therefore, a cylindrical transducer embedded in a composite host plate has been considered. The electrical impedance of the transducer alone in vacuum and then of the embedded transducer, which allows the identification of the resonance modes, have been obtained by the finite-element method (FEM). Moreover, the displacement fields in the plate, which allow the identification of the type of Lamb waves generated, have been computed at the resonance frequencies. Comparison between the FEM results and the Lamb waves dispersion curves of the host material are in good agreement. Experimental results (electrical impedance, frequency response, and phase velocities) concerning a composite plate specimen containing the same piezoelectric transducer inside it will be shown. A good agreement is generally obtained between numerical and experimental results.

11:15

2aEA9. Signal processing for detection of defect using Lamb waves. Pierrick Blanquet, Thierry Demol, Emmanuel Moulin, and Christophe Delebarre (IEMN DOAE, Univ. de Valenciennes, BP311 59304, Valenciennes Cedex, France)

Conventional ultrasonic inspection of large structures is very time consuming due to the surface scanning necessity. Since they can propagate over long distances, Lamb waves are an attractive solution to this problem. Here is shown the feasibility of using Lamb wave generation using a transducer embedded inside the material and two methods are given to determine the presence of a defect along a propagation line. In a first step, the embedded transducer is used to generate and receive Lamb waves. Then, a method based on the spatial Fourier transform is used to measure their phase velocities. The experimental measurements are in good agreement with a numerical phase velocity computation. In a second step, the

determination of the presence of a defect along a propagation line is shown. One method consists in measuring the attenuation of Lamb waves during the propagation. Another one is based on the Hilbert transform, which can be used to determine mode conversions due to the damage. In both cases, the results show the feasibility of determining the presence of a defect.

11:30

2aEA10. Interaction of acoustical waves in concretes with cracks. Igor N. Didenkulov, Alexander E. Ekimov, Vyacheslav V. Kazakov, Alexander M. Sutin (Inst. of Appl. Phys., 46 Ulyanov St., Nizhny Novgorod, 603600, Russia), Paul A. Johnson (Los Alamos Natl. Lab., Los Alamos, NM 87545), Igor N. Budnikov, Boris V. Bagryanov, and Mikhail B. Denisov (Russian Federal Nuclear Ctr., Nizhny Novgorod Region, 607190, Russia)

Two separate experiments conducted with concrete samples containing cracks illustrate that acoustical methods have promise in damage detection. Samples of concrete were 2.6 m×19 cm×12 cm and 1.7 m×20 cm×20 cm for the first and the second experiments, correspondingly. Cracks were located near sample centers and had dimensions of sample sections. In one experiment, high-frequency (5.3-kHz) longitudinal or rotational waves were modulated by low frequency (3 to 200-Hz) flexural vibrations. The relative amplitude of the modulation signal was about −20 dB. When the crack was filled by water, the modulation amplitude decreased in amplitude by approximately 3–4 dB. This result is expected because fluid should diminish the nonlinearity on the crack contact. In an uncracked sample, no modulation should be observed. In the second experiment, small explosive sources (less than 100 kg/cm^2) were used to study the interaction of a large amplitude, broad-band frequency pulse, with a weak impulse from small explosive source. It was first observed that the weak pulse did not propagate through the crack. The weak signal passed through the crack when a stronger explosion pulse was simultaneously applied to the bulk, transiently closing the crack. The results are two of many experiments that indicate linear and nonlinear wave methods may be applied to characterize damage in concrete. [Work supported by INCAS, Nizhny Novgorod and by RFBR Grant 97-02-17524.]

TUESDAY MORNING, 2 DECEMBER 1997 COUNCIL ROOM, 8:45 A.M. TO 12:00 NOON

Session 2aMU

Musical Acoustics: General

R. Dean Ayers, Chair
Department of Physics and Astronomy, California State University, Long Beach, California 90840

Contributed Papers

8:45

2aMU1. Mesa di voce: An investigation of the symmetry of crescendo and decrescendo in singing. Ingo Titze (Dept. of Speech Pathol. and Audiol. & Natl. Ctr. for Voice and Speech, Univ. of Iowa, Iowa City, IA 52242) and Russel Long (Denver Ctr. for the Performing Arts, Denver, CO 80204)

Messa di voce (to place the voice) is an often used exercise by professional classical singers. It is a crescendo followed by a decrescendo on a constant pitch and vowel. Difficulty is often encountered in maintaining constant voice quality (register) and a consistent and gradual loudness change, especially at high pitches. Six well-trained singers were studied, three males and three females. Measurements consisted of sound pressure level, electroglottography, oral airflow, subglottal pressure, lung volume change, and ribcage and abdomen movement. Results indicate that symmetry between crescendo and decrescendo portions is effected by variations in subglottal pressure and peak acoustic airflow, and to a lesser extent, by abdominal movement. Subjects also differed greatly in their soft–loud dynamic ranges. [Work supported by NIDCD, Grant No. P60 DC00976.]

9:00

2aMU2. Experiments on a recorder with a movable, expandable tone hole. Kay C. Kunes and Donald E. Hall (Phys. Dept., California State Univ., 6000 J St., Sacramento, CA 95819)

The pitch produced by a flute-type instrument with a single tone hole depends on the position and size of that hole. Comparisons of theory with reality would be easiest if that tone hole had continuously variable size and position, so that no other holes need be present to complicate matters. A scheme for having such a tone hole has recently been presented [D. Hall and J. Schreiner, Proc. Intl. Soc. Musical Acoustics, Edinburgh (1977)]. A recorder head joint attached to a slotted cylindrical tube fitted with a sleeve with a wedge-shaped hole allows ranges of approximately 5 cm in hole position, a factor of 5 in hole size, and 2.5 semitones in pitch. This makes possible relatively simple experiments for (1) comparing the effect of hole size and position upon fundamental frequency with approximate theories and (2) studying the effect of hole size and position upon departure from harmonicity of the first two natural modes of the instrument. Results of such experiments will be presented.

9:15

2aMU3. Modeling free reed behavior using calculated reed admittance. James P. Cottingham and Casey A. Fetzer (Phys. Dept., Coe College, Cedar Rapids, IA 52402)

For musical wind instruments driven by a reed mechanism, Fletcher has provided an analysis which shows that a necessary condition for the sounding of an instrument is that the acoustic admittance of the reed has a real part, which is negative and greater in magnitude than the real part of the pipe admittance [N. H. Fletcher, Acustica **43**, 63–72 (1979)]. The analysis also shows that inward-striking (closing) and outward-striking (opening) reeds will exhibit different behavior. Free reeds, used in instruments such as the reed organ, harmonica, accordion, and the Asian free reed mouth organs, can exhibit behavior characteristic of either opening or closing reeds, depending on the details of the configuration. Using reed parameters determined by experiment, the reed admittance as a function of frequency has been calculated for free reeds and used to make quantitative predictions. For a variety of free-reed parameters, the frequency of vibration as a function of blowing pressure has been calculated and compared with experimental results. Cases are considered in which a reed is coupled to a pipe resonator as well as cases in which the reed is not coupled to a resonator.

9:30

2aMU4. Vibrational modes of clarinet reeds. Ian M. Lindevald and Jason Gower (Div. of Sci./Phys., Truman State Univ., Kirksville, MO 63501)

The role of the lowest frequency reed resonance in sound production by single reed instruments has been well established [S. C. Thompson, J. Acoust. Soc. Am. **66**, 1299–1307 (1979)]. However, the acoustical significance of the reed's higher vibrational modes remains an open question. Recently, several higher frequency modes have been identified for a clarinet reed [P. L. Hoekje and G. Matthew Roberts, J. Acoust. Soc. Am. **99**, 2462(A) (1996)]. The present study examines several reeds and extends Hoekje's results to higher frequencies. Holographic interferometry was used to locate the frequencies and to image the vibrational patterns of higher vibrational modes of three nominally equivalent commercial clarinet reeds. Dry reeds were clamped to a specially designed reed holder allowing the reeds to vibrate freely when driven by an acoustic sound field. In all, six or seven distinct modes with frequencies below 7.5 kHz were located on each of the three reeds. Four of the modes (including the fundamental) appear to be similar to the one-dimensional modes of a cantilevered beam. The other three modes show two-dimensional vibrations indicating twisting motion. Variations in modal frequencies among the three reeds studied range from 3%–12% over the different modes.

9:45

2aMU5. How does a cone mimic a doubly closed pipe? R. Dean Ayers (Dept. of Phys. and Astron., California State Univ., Long Beach, 1250 Bellflower Blvd., Long Beach, CA 90840, rdayers@csulb.edu)

An idealized conical bore that is complete to its apex has been used as the basic model for several wind instruments with tone holes, such as oboe, bassoon, saxophone, and cornetto. The useful playing frequencies for these relatively short instruments include those near the first peak in $|Z|$. A uniform pipe that is closed at both ends has been proposed as the basic bore shape for the brasses, including even their strongly conical representatives, such as tube and cavalry bugle [R. D. Ayers, J. Acoust. Soc. Am. **98**, 81–87 (1995)]. The relationship should be considered one of mimicry, since there are major discrepancies at low frequencies, and these extend to the first (anti)resonance or higher. The nature of that relationship and the way in which it is achieved are perhaps most easily understood by examining a conical frustum, with or without a lumped-element mouthpiece, from this new perspective. Fairly simple behaviors in the time domain provide the explanation for the mimicry in the frequency domain.

10:00

2aMU6. Harmonic alignment of impedance peaks in brass instruments. R. Dean Ayers (Dept. of Phys. and Astron., California State Univ., Long Beach, 1250 Bellflower Blvd., Long Beach, CA 90840, rdayers@csulb.edu)

Given that a nearly harmonic alignment is a desirable condition, to what extent is it achieved in a particular brass instrument, and how do the bell and the input segment cooperate to bring it about? Answers to these questions that are easily read by human beings are not to be found in a plot of $|Z|$, or any aspect of the plane-wave reflection coefficient R at the input end of the air column. Instead, the bell and the input segment are each examined from inside the central segment and represented in terms of the effective length to a closed end, which has a simple relationship to the phase of the (internal) reflection coefficient. The sum of those two lengths should be essentially independent of frequency for harmonic alignment. This criterion is very easy to evaluate on a graph, and it applies simultaneously to all lengths of the central segment, hence all fingerings or slide positions. Minor perturbations on this relationship due to processing by the input segment and viscothermal effects on the speed of sound in the central segment will be discussed.

10:15–10:30 Break

10:30

2aMU7. Relevant acoustical cues in the identification of Western orchestral instrument tones. John M. Hajda (Music Percept. and Cognit. and Musical Acoust. Lab., Dept. of Ethnomusicol., UCLA, Box 951657, Los Angeles, CA 90095-1657, e-mail: jhajda@ucla. edu)

The purpose of this research is to establish perceptually relevant operational definitions of the acoustical segments of isolated musical instrument tones. Previous research has yielded conflicting conclusions for continuant (nonpercussive) tones regarding the relative salience of the attack and steady-state portions. This study also examines the perception of impulse (percussive) instrument tones. A new model for the partitioning of continuant tones, the ''amplitude/centroid trajectory,'' divides the signal into four segments: The attack, attack/steady-state transition, steady state, and decay. The divisions are based on the direction of and relationship among the global rms amplitude and spectral centroid. Other operational definitions of signal partitions and reverse playback were included as editing treatments. Single isolated tones were used from six impulse and six continuant instruments. Nine subjects participated in a 246-item forced-choice identification experiment. It was determined that, for sustained continuant tones, the steady state alone provides sufficient and necessary information for the identification of the instrument. The coevolution of the amplitude and spectral centroid seem important, but the direction does not. A confounding of envelope with identification for staccato tones was found. For impulse tones, both the direction *and* coevolution of global components are necessary.

10:45

2aMU8. Perception of sequences of octave-related complexes in relation to the tritone paradox. Frank Ragozzine (Dept. of Psych., Univ. of California, San Diego, Mail Code 0109, 9500 Gilman Dr., La Jolla, CA 92093)

Shepard [J. Acoust. Soc. Am. **36**, 2346–2353 (1964)] has shown that ordered pairs of octave-related complexes are perceived in accordance with proximity. However, in his experiment, proximity-based judgments gradually decreased as intervals approached a tritone. With tritones, Deutsch [Music Percept. **8**, 335–347 (1991)] found that listeners perceive half of the pitch-class circle as higher in pitch, and the opposite half as lower. The half heard as higher differs among subjects (the tritone paradox). Deutsch theorized that listeners compare these tritones to a circular pitch-class template, whose orientation with respect to height differs among listeners. The present study examines the perception of octave-related complexes that form intervals of P5/P4 or m3/M6. It was hypothesized that listeners may sometimes use a pitch-class template, rather than proximity, with intervals other than a tritone. Template orientations were determined using Deutsch's paradigm. Subjects were presented with sequences of tones that traversed the circle of fifths. Perception of these tones was closely related to each subject's pitch-class template, and not to proximity. Similar, though less pronounced, results were found in a second experiment using an ordered series of m3/M6.

11:00

2aMU9. Use of measured data in the physical modeling of the classical guitar. Ricardo R. Boullosa and Felipe Orduña-Bustamante (Centro de Instrumentos UNAM, Circuito Exterior CU, México DF, CP 04510, Mexico)

A simple physical model consisting of a delay-network representation of the string and a radiation filter (which transduces bridge motion into sound pressure radiated at a distance) is used to synthesize the sound of the classical guitar. Parameters for the sound generation algorithm are obtained from mechanical and acoustical transfer functions measured in a guitar. These correspond to mechanical response at the bridge, or to acoustical response at a given point in the sound field, and involve mechanical excitation in the bridge at different string positions. The variation from string to string observed in these measurements, especially at high fre-

quencies, suggests the need for individual characterization of each string in the physical model of the guitar. Other factors, such as stiffness (inharmonicity), internal string damping, and nonlinear string motion are also discussed as desirable additional features of more realistic sound generation algorithms.

11:15

2aMU10. Subjective evaluation of classical guitars. Felipe Orduña-Bustamante and Ricardo R. Boullosa (Centro de Instrumentos UNAM, Circuito Exterior CU, México DF, CP 04510, Mexico)

An analysis is presented of two methods for the subjective evaluation of a set of classical guitars. One is based on ''blind'' listening tests by a general audience, and another is based on direct inspection and playing by guitar teachers and students. Subjective experiments reveal that the two types of tests produce essentially the same sorting of the guitars by subjective quality. The listening test, based on pairwise comparisons of guitars playing the same musical fragment, is judged by the subjects to be a difficult task. This is also reflected in the results. Additionally, an interesting effect is consistently observed in which a higher proportion of votes is obtained by the second guitar played in each pair, despite the fact that listening to each pair of guitars lasts less than about 40 s. This seems to suggest a certain natural limit in the short-term memory abilities of the subjects in recalling the first guitar. A data processing technique is described that produces unbiased results from this type of test. The playing test, on the other hand, is judged by the players to be a comparatively easy task and the results are also more easily interpreted.

11:30

2aMU11. On the use of the instantaneous frequency to understand phenomena in polyphonic piano sounds. Lucile Rossi and Gerard Girolami (Univ. of Corsica, Quartier Grosseti, 20250 Corte, France, lrossi@univ-corse.fr)

The extraction of amplitudes and frequencies of partials from musical tones have been developed over the past ten years in order to carry out statistical studies on spectral parameters [Ando, J. Acoust. Soc. Am. **94**, 37–45 (1993)] and/or to measure the frequency ratios of spectral components [Brown, J. Acoust. Soc. Am. **99**, 1210–1218 (1996)]. This paper presents a study of the temporal evolution of the amplitude and frequency of piano sounds' partials in order to take into account the eventual frequential fluctuations in the identification method of polyphonic piano sounds developed by the authors [Rossi and Girolami, J. Acoust. Soc. Am. **100**, 2842 (1996)]. The notion of analytical signal and instantaneous frequency is used to find the characteristics of beats appearing in the evolution of the amplitude of single notes and single notes' partials. In the case of polyphonic piano sounds, pertubations in the temporal evolution of frequency and amplitude of a given partial, due to the presence of a close partial belonging to another note, are presented.

11:45

2aMU12. Heterogeneity of maple wood for musical instrument making. Emil Rajčan and Stanislav Urgela (Dept. of Phys. and Appl. Mech., Tech. Univ. Zvolen, Masarykova 24, SK-960 53 Zvolen, Slovakia)

Maple wood for back plates of bowed string instruments is more and more rare. The measuring of its properties is then, very topical. This paper is concerned with physicoacoustical characteristics such as density, Young's modulus of elasticity, and the acoustic constant, and their distribution along the radial direction in the trunk of a tree. The study was performed with the nondestructive Chladni method on plates. A bending beam vibrational mode along the grain was visualized and the pictures indicate interesting shapes caused by the heterogeneity of the wood and internal stress. Then, the plates were cut to sticks and the same mode was visualized. Also here, the distortion of nodal lines, when compared with homogeneous material, has been shown. The elastic characteristics were calculated from the natural frequencies of the modes. Such testing has been proven as useful for the quality determination of maple wood as a material for musical instrument making. [Work supported by Slovak Grant Scientific Agency.]

TUESDAY MORNING, 2 DECEMBER 1997 CABINET ROOM, 8:30 TO 11:30 A.M.

Session 2aPA

Physical Acoustics: Radiation, Propagation and Scattering

James P. Chambers, Chair

National Center for Physical Acoustics, University of Mississippi, Coliseum Drive, University, Mississippi 38677

Contributed Papers

8:30

2aPA1. Equivalent phased array methods to predict acoustic fields of planar and focused ultrasound transducers. Eduardo G. Moros, Xiaobing Fan, and William L. Straube (Washington Univ., 4511 Forest Park Ave., Ste. 200, St. Louis, MO 63108)

Acoustic fields generated by single-source ultrasonic transducers were predicted accurately by combining phased array theory and the Rayleigh–Sommerfeld diffraction integral. Two methods were developed and experimentally tested. The first is the equivalent phased array method (EPAM), which treats any small transducer as a phased array composed of many square elements [Moros *et al.*, J. Acoust. Soc. Am. **101**, 1734–1741 (1997)]. The second method, called CREPAM for the concentric-ring equivalent phased array method, was developed for larger axisymmetric transducers. The goal of both methods is to determine the excitation source of the equivalent phased array employing an inverse technique and data measured at a plane in the near field and perpendicular to the transducer's axis. The determined excitation source and equivalent phase array can then be used to compute the acoustic field from the measurement plane to the far field. It was demonstrated that both methods accurately predict the location of major grating lobes and general patterns of the near and far pressure fields for several planar and focused transducers. These methods should be useful in diagnostic and therapeutic ultrasound applications where transducer-specific field characteristics are important. [Supported by NCI Grant R29-CA63121.]

8:45

2aPA2. Ultrasonic properties of aluminosilicate glass. J. H. So (Dept. of Phys. and Astronomy, CMSS Program, Ohio Univ., Athens, OH 45701, jhso@helios.phy.ohiou.edu), D. H. Green, and S. S. Yun (Ohio Univ., Athens, OH 45701)

An ultrasonic study of attenuation and sound velocity was carried out in aluminosilicate glasses including annealed samples at 800 and 900 °C for 10 h as functions of temperature and frequency with the pulse–echo method. The temperature and frequency ranges employed here were 100–300 K and 5–50 MHz. The original sample has the largest attenuation coefficient in the temperature range compared to that of samples annealed. In all samples, the longitudinal and shear-wave velocities decrease linearly with temperature. Young's, shear, and bulk moduli decrease linearly with temperature also. In this temperature range, Lame's parameter, Poisson's ratio, and the velocity anisotropy factor were determined using velocity data. The ultrasonic attenuation coefficient increases linearly with frequency and monotonically with decreasing temperature and is independent of pulse amplitudes. These results imply a linear loss mechanism producing a frequency-independent Q.

9:00

2aPA3. Coherence calculations for propagation through anisotropic, turbulent velocity fluctuations. D. Keith Wilson (U.S. Army Res. Lab., 2800 Powder Mill Rd., Adelphi, MD 20783, dkwilson@arl.mil)

Previous treatments of the mutual coherence function (MCF, describing the coherence between the signals received by two sensors as a function of their spatial separation) have been based on isotropic turbulence models. However, acoustic waves propagating through the atmosphere are strongly affected by a large-scale turbulence structure, which is anisotropic by nature. In this paper, the MCF is calculated using several rigorous, anisotropic models for turbulent velocity fluctuations in the neutral atmospheric shear layer: a generalized Gaussian model; a generalized von Kármán model; the Kristensen *et al.* model [Bound. Layer Meteor. **47**, 149–193 (1989)]; and Mann's rapid-distortion, shear model [J. Fluid Mech. **273**, 141–168 (1994)]. Although it is possible to obtain analytical results for the MCF for the isotropic Gaussian model and von Kármán models, one can no longer do so when these models are extended in a suitable fashion to the anisotropic case. However, it is not difficult to obtain numerical results. By comparing the anisotropic models to their isotropic versions, it is shown that anisotropy can significantly affect coherence, even when the spacing of the sensors is small (in comparison to the height from the ground).

9:15

2aPA4. Confirmation of the effective coefficient for scattering by turbulent velocity fluctuations. Makeda J. Smith, D. Keith Wilson (U.S. Army Res. Lab., 2800 Powder Mill Rd., Adelphi, MD 20783), Michael Heyd, David I. Havelock (Natl. Res. Council, Ottawa, ON K1A 0R6, Canada), Harry J. Auvermann, and John M. Noble (U.S. Army Res. Lab., Adelphi, MD 20783)

The effective structure-function parameter for scattering by atmospheric, turbulent velocity fluctuations has normally been assumed to be $C_{\text{eff}}^2 = 4C_V^2/c_0^2$ However, a new derivation by V. E. Ostashev [Waves Random Media **4**, 403–428 (1994)], which takes into account the vectorial nature of the wind velocity field, suggests that $C_{\text{eff}}^2 = 22C_V^2/3c_0^2$. An experiment was designed to determine the correct value of the coefficient. Amplitude variances were monitored for several discrete frequencies between 380 and 3500 Hz, at distances up to 675 m. Cup and hot-wire anemometers were used to determine C_V^2. A theory for scattering by inertial-subrange turbulence was then used to calculate the C_{eff}^2 coefficient from the amplitude variance and C_V^2. Although there is some tendency in favor of the 22/3 coefficient, the results cannot be considered conclusive. The main difficulty appears to be obtaining a sufficiently accurate measurement of C_V^2.

9:30

2aPA5. Strategies for computing sound propagation through realistic three-dimensional turbulence. Xiao Di and Kenneth E. Gilbert (Appl. Res. Lab. and the Grad. Program in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804)

For 3-D sound propagation calculations, if computation time were not an issue, it would be straightforward to extend 2-D parabolic equation (PE) models to three dimensions. However, for any 3-D propagation model, computation time is a crucial factor on any personal computer presently available. Hence, in this paper, some computational approaches are discussed that allow one to treat realistic 3-D turbulence in an approximate but physically reasonable way. The computational methods preserve the 3-D statistical properties of the turbulence without generating an unacceptable computational burden. The proposed approaches are outlined and numerical examples are given for sound propagation through meteorologically realistic turbulence fields. [Work supported by DARPA and the Pennsylvania State University Applied Research Laboratory.]

9:45

2aPA6. First symbiosis between rigorous ray tracing and the parabolic approximation. Edward R. Floyd (10 Jamaica Village Rd., Coronado, CA 92118-3208)

Rigorous ray tracing originates from the trajectory representation of quantum mechanics while the (Leontovich–Fock) parabolic equation was originally applied to electromagnetic propagation. Rigorous ray tracing (and quantum trajectory representation) have had theoretical success while the parabolic approximation has been a computational success. The parabolic approximation inspires a similar equivalent approximation to the generalized Hamilton–Jacobi equation for rigorous ray tracing. For separable coordinates, the resulting approximations render the same approximate wavefunction. In return, the rigorous ray leads to a parabolic approximation incorporating reflection. The trajectory representation has rendered a synthesized single wavefunction that includes an incident wave of coefficient α and a reflected wave of coefficient β. This synthesized wavefunction can generate a new alternative parabolic equation that incorporates reflection, which for small reflection ($\beta \ll \alpha$) becomes $F_{zz} + i2k[1-(2\beta/\alpha)\cos(2kx)]\exp[i(2\beta/\alpha)\sin(2kx)]F_x + (\kappa^2 - k^2)F = 0$.

10:00–10:15 Break

10:15

2aPA7. Backscattering enhancements due to retroreflection of leaky Rayleigh waves at the truncation of a finite-solid cylinder: Measurements and test of a theory. Karen Gipson and Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814, gipson@wsunix.wsu.edu)

Leaky surface waves (such as Rayleigh waves) once launched on an elastic object will undergo partial reflection at the truncations of the object. Under certain conditions, such reflections can radiate locally flat acoustic wavefronts with a reversal of the acoustic-wave vector, and thus a large backscattered signal is produced. In the present experiments, a solid stainless-steel cylinder was ensonified through an angular region, which included the conditions for launching meridional and helical Rayleigh waves, and the resulting backscattering enhancements are consistent with hypothesized reflection processes. For the enhancement associated with a meridional ray, the observed peak magnitude of the retroreflected signal is in general agreement with approximate theoretical predictions from a convolution formulation [P. L. Marston, J. Acoust. Soc. Am. **102**, 358–369 (1997)]. Helical ray contributions are also observed, but these are generally weaker at ultrasonic frequencies.

10:30

2aPA8. Bravais effective refractive index for tilted plastic cylinders and the caustic-merging transition in the meridional plane: Demonstration of an optical analogy. Philip L. Marston, Catherine M. Mount, and David B. Thiessen (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

Some solid scatterers in water such as plastic cylinders do not support leaky Rayleigh waves because the shear wave velocity of the solid is less than the speed of sound in water. The high-frequency scattering from such objects is generally weak. In that case the ray diagram for transmitted shear waves is analogous to the refraction of light by an object in air. For a tilted cylindrical object, the projection of rays on a base plane may be determined by ignoring the tilt by replacing the true refractive index n by the Bravais value $(n^2-\sin^2\gamma)^{1/2}/\cos\gamma$ where γ is the tilt. For a circular cylinder, when the Bravais value reaches 2, the rainbow caustics merge in the meridional plane defined by the incident wave vector and the cylinder's axis [C. M. Mount and P. L. Marston, in *Light and Color in the Open Air* (OSA, Washington, DC, 1997), pp. 14–16]. The merged caustics greatly enhance the scattering because of the far-field focusing of a cusp point. The analysis is also relevant to light scattering by icicles and to the scattering of sound by certain liquid-filled cylinders. [Work supported by the Office of Naval Research.]

10:45

2aPA9. Backscattering enhancement for tilted plastic cylinders in water due to the caustic-merging transition: Ultrasonic observations. Florian J. Blonigen and Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

Typical plastics do not support leaky Rayleigh waves in water because the shear velocity is less than the speed of sound in water. Consequently, backscattering from plastic cylinders is ordinarily small. If the shear-wave attenuation is sufficiently small, there is, however, a mechanism for enhanced backscattering by tilted circular cylinders. At a certain tilt angle, rainbow caustics merge in the meridional plane defined by the incident wave vector and the cylinder's axis. The merged caustics are associated with rays internally reflected off an end of the cylinder but are analogous to a merging of caustics observed [C. M. Mount and P. L. Marston, *Light and Color in the Open Air* (OSA, Washington, DC, 1997), pp. 14–16] in optical experiments with long plastic fibers. Ultrasonic observations of the backscattering enhancement are reported for a plastic cylinder in water with $ka=25$. The enhancement mechanism should also apply to liquid cylinders in a thin shell if the speed of sound for the liquid is less than that of water. [Work supported by the Office of Naval Research.]

11:00

2aPA10. Scattering of pulse signals from fluid-filled elastic shells. Natalia A. Sidorovskaia (Phys. Dept., Univ. of New Orleans, New Orleans, LA 70148) and Michael F. Werby (Code 7181, NRL, Stennis Space Center, MS 39529)

The study of the sound scattering from fluid-filled elastic shells has shown a formation of the complicated return structures when scattering from such objects takes place. Preliminary work has suggested that the apparent complicated structure may be resolved in terms of a superposition of standing waves at the interface. Indeed, now, when a new acoustic background has been developed, the coherent vibrational effects may be resolved into patterns that are likely connected with the curved surface analog of Sholte waves on flat plates that are fluid loaded on both sides. Theoretical partial wave studies suggest this, but in order to ''see'' it in experiment, one must resort to pulse studies and gating methods. This hypothesis is tested by simulating experimental data and employing signal-processing techniques to determine if the effect can be observed in real experiments. [Work sponsored by ONR, the Naval Research Laboratory, and the University of New Orleans.]

11:15

2aPA11. FDTD boundary treatment for scattering from penetrable objects. Shuozhong Wang, Tong Wang (Dept. of Electron. Info. Eng., Shanghai Univ., Shanghai 200072, China), and Rongqing Wang (760 Res. Inst., Dalian 116013, China)

The finite-difference time-domain (FDTD) formulation [S. Wang, J. Acoust. Soc. Am. **99**, 1924–1931 (1996)] and the least-squares absorbing boundary conditions [S. Wang, Chin. J. Acoust. **16**, 121–134 (1997)] have been successfully applied to deal with sound scattering from ideally soft and ideally rigid objects with arbitrary shapes. To expand the scope of application to more realistic cases, an effective boundary treatment for interfaces penetrable to sound waves is introduced. The FDTD boundary condition expressions take a form similar to that in continuous and uniform media, except for a modification to the material-related constants. This is achieved by using a suitable averaging scheme based on a second-order discretization of the basic acoustic equations on boundaries. The proposed method is consistent with the ideal cases at both soft and rigid ends, giving a unified approach to the solution of scattering from nonelastic objects. Numerical experiments on scattering from various objects submerged in water are presented, showing a good agreement with theoretical solutions and previously obtained results. The object materials used in the computation cover a wide range of acoustic impedance from very soft to very rigid. [Work supported by NNSFC.]

Session 2aSA

Structural Acoustics and Vibration: Computational Vibroacoustics I

Noureddine Atalla, Chair
G.A.U.S., Mechanical Engineering, University de Sherbrooke, Sherbrooke, Quebec J1K 2R1, Canada

Chair's Introduction—8:40

Invited Papers

8:45

2aSA1. Capabilities and limitations of boundary- and finite-element methods for solving noise and vibration problems. M. A. Hamdi (Univ. of Technol. of Compiegne, 602200, Compiegne, France, hamdi@straco.fr) and L. Mebarek (Straco, 60471, Compiegne Cedex, France)

Over the last decades the boundary- and finite-element methods (BFEM) have been used in the solution of practical mechanical engineering problems. It is, therefore, appropriate at this time to examine the capabilities and the limitations of these numerical methods to predict, analyze, and solve real-life vibroacoustic problems encountered in several industries such as aerospace, automotive, naval, and railroad. Due to the rapid growth of computer power, several computer-aided engineering tools, based on these BFEM techniques, have been developed and are now commercially available. The purpose of this paper is to review the current state of development of BFEM techniques, and to examine their ability to predict and solve noise and vibration problems. The finite-element method (FEM), will be first analyzed for interior fluid–structure interaction problems, with an extension for exterior radiation problems using the infinite-element method (IEM). Then, the boundary-element method (BEM), will be analyzed for external interaction problems, where the fluid domain is infinite. It will be shown that the combination of these two methods (BFEM) provides a very powerful tool for solving a wide class of vibroacoustic problems, taking advantage of advanced numerical algorithms implemented on modern high-performance computing platforms.

9:15

2aSA2. Analysis of fluid-driven panels using the energy finite-element method. Fei Han, Luc G. Mongeau, and Robert J. Bernhard (Ray W. Herrick Labs., School of Mech. Eng., Purdue Univ., West Lafayette, IN 47907-1077)

Energy flow analysis methods have been developed to predict the approximate response of built-up structural-acoustic systems at high frequency. The governing equations of energy flow analysis can be formulated into a finite-element approximation, referred to as the energy finite-element method. In recent investigations, a method has been developed to calculate the power input for energy flow analysis of systems with random, distributed loads. The method uses cross spectra pressure information. The method is intended to allow energy flow analysis and energy finite-element analysis of flow excited panels. Both analytical and experimental verification studies have been done to test the approach for panels excited with distributed pressure loads. The predicted results are excellent approximations of the response.

9:45

2aSA3. Overview of multipole-based acoustic infinite elements. David S. Burnett (Bell Labs., Lucent Technol., Rm. 15F-329, 67 Whippany Rd., Whippany, NJ 07981, dsburnett@lucent.com)

There is a rapidly growing interest in using infinite elements based on multipole expansions for modeling acoustic wave propagation in unbounded domains, either time harmonic or transient. The technology has exhibited many advantages, foremost among them being a very high computational efficiency, convergence and, therefore, unlimited accuracy throughout the entire fluid domain, broadband capability, and a conventional finite-element software implementation. The latter enables these infinite elements, along with conventional finite-size acoustic elements and coupling elements, to be easily added to standard structural element libraries to convert structural codes to structural acoustic codes for fast, accurate modeling of scattering and/or radiation problems in exterior regions. The talk will cover: basic concepts of multipoles, multipole expansions, and the related infinite elements; past contributions to the field; current research, including the author's; and an assessment of the most promising direction for this technology.

10:15

2aSA4. Exploring mid-frequencies with a ''hybrid statistical energy analysis'' method. Paul G. Bremner (Vibro-Acoust. Sci., Inc., 5355 Mira Sorrento Pl., #100, San Diego, CA 92121) and Robin S. Langley (Univ. of Southampton, SO17 1BJ, United Kingdom)

The dynamical behavior of a complex structure at midfrequencies poses special analytical modeling challenges. Traditionally, finite-element analysis (FEA) has been viewed as an analysis method for low-frequency vibrations, while statistical energy analysis (SEA) has been used to analyze high-frequency vibrations. Ongoing work funded by the RESOUND Consortium is directed at bridging the gap between these two methods to produce an analysis technique which is applicable across the whole frequency range, including the problematic ''midfrequency'' region. A novel wave-number partitioning scheme has been developed under which the

system dynamic response is represented as a combination of finite-element ''global modes'' and SEA subsystem ''local modes.'' It has been found that the interaction between these two sets of modes encompasses many of the features of ''fuzzy structure theory,'' an area of great current interest in structural dynamics. In many design situations a (relatively) course mesh finite-element model of a vehicle is developed to study static and low-frequency dynamical behavior. Such a model can be incorporated into the proposed method without the need for labor intensive remodeling. This paper will describe initial applications of the method that was introduced by the authors in a previous abstract [J. Acoust. Soc. Am. **100**, 2754(A) (1996)]. In particular, it will present the results of some experimental validation of the method applied to midfrequency vibration response of a truss-frame structure.

10:45–11:00 Break

Contributed Papers

11:00

2aSA5. Helmholtz equation–least-squares method for reconstructing acoustic pressure fields. Zhaoxi Wang (NVH Lab., Chrysler Corp., 800 Chrysler Dr., Auburn Hills, MI 48326) and Sean F. Wu (Wayne State Univ., Detroit, MI 48202)

A method using spherical wave expansion theory to reconstruct an acoustic pressure field from a vibrating object is developed. The radiated acoustic pressures are obtained by means of an expansion of independent functions generated by the Gram–Schmidt orthonormalization with respect to the particular solutions to the Helmholtz equation on the vibrating surface under consideration. The coefficients associated with these independent functions are determined by requiring the assumed form of the solution to satisfy the pressure boundary condition at the measurement points. The errors incurred in this process are minimized by the least-squares method. Once these coefficients are specified, the acoustic pressure at any point, including the source surface, is completely determined. In this paper, this method is used to reconstruct the surface acoustic pressures based on the measured acoustic pressure signals in the field. It is shown that this method can be applied to both separable and nonseparable geometries, and the solutions thus obtained are unique. However, the convergence may be worse for an elongated object subject to high-frequency excitations. [Work supported by the Institute for Manufacturing Research at Wayne State University.]

11:15

2aSA6. Reconstruction of acoustic pressure fields inside a vibrating cavity. Sean F. Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202) and Jingyou Yu (Wayne State Univ., Detroit, MI 48202)

The Helmholtz equation–least-squares (HELS) method developed by Wang and Wu [J. Acoust. Soc. Am. (to be published)] is extended to reconstruction of acoustic pressure fields inside a vibrating cavity. In this method, the acoustic pressures are reconstructed through an expansion of acoustic modes, or a set of independent functions generated via an orthonormalization with respect to the particular solutions to the Helmholtz equation on the particular surface under consideration. The coefficients associated with these acoustic modes are determined by requiring the assumed form solution to satisfy the pressure boundary condition at the measurement points. The errors incurred in this process are minimized by the least-squares method. Numerical examples of partially vibrating spheres and cylinders with various half-length to radius ratios and dimensionless frequencies are demonstrated. The results thus obtained are compared with the numerical solutions obtained by using a direct boundary-element method (BEM). Results show that the HELS method enables one to reconstruct the radiated acoustic pressure fields much more effectively and accurately than BEM does. [Work supported by the Institute for Manufacturing Research at Wayne State University.]

11:30

2aSA7. An alternative formulation for predicting sound radiation from a vibrating object. Sean F. Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202) and Qiang Hu (Wayne State Univ., Detroit, MI 48202)

An alternative formulation is derived for predicting acoustic radiation from a vibrating object in an unbounded fluid medium. The radiated acoustic pressure is shown to be expressible as a surface integral of the particle velocity, which is determinable by using a nonintrusive laser velocimeter. This approach is in contrast with the Kirchhoff integral formulation which requires the knowledge of both the normal component of the surface velocity and the surface acoustic pressure prior to predicting of the radiated acoustic pressure. Solutions thus obtained are unique. Moreover, the efficiency of the numerical computations is high because the surface integration can be readily implemented numerically by using standard Gaussian quadratures. Such a formulation can be readily extended to acoustic scattering problems and acoustic radiation in an interior region. This alternative formulation may be desirable to analyze the acoustic and vibration responses of a lightweight, a flexible, or a structure under an adverse environment for which a nonintrusive laser measurement technique must be used. Validations of this alternative formulation are demonstrated both analytically and numerically for vibrating spheres and right circular cylinders. [Work supported by NSF.]

11:45

2aSA8. Structural sensitivity analysis in structure-fluid interaction problems. Steven R. Hahn and Aldo A. Ferri (Georgia Tech, Mech. Eng. Dept., MS0405, Atlanta, GA 30033)

Using a finite-element formulation with infinite elements to approximate the external fluid, well-developed structural sensitivity analysis techniques can be applied to acoustic radiation and scattering problems. Changes to a nominal structure are introduced through a structural perturbation matrix, $\epsilon\Delta\mathbf{S}$, which modifies the system matrix, $\mathbf{S}$, of the nominal system. The perturbed surface pressures and normal velocities are expressed as a binomial series in $\Delta\mathbf{S}$. The convergence criteria for the series is examined for structural-acoustic systems. A perturbation consisting of a change in the Young's modulus is examined for two numerical examples. First, a one-dimensional problem of plane-wave scattering from a viscoelastic layer is examined. The results of the 1-D problem are then extended to the multidimensional problem of plane-wave scattering from a cylindrical shell. The perturbations consist of varying the Young's modulus of the layer and the shell. It is shown that the convergence of the binomial series is strongly linked to the amount of damping in the nominal system, but not to the damping in the perturbation. Furthermore, the use of a truncated series is shown to be a valid approximation in many cases, even when the series does not converge.

Session 2aSC

Speech Communication: Perception and Production of Vowels and Consonants (Poster Session)

Terrance M. Nearey, Chair
Department of Linguistics, University of Alberta, 4-32 Assiniboia Hall, Edmonton, Alberta T6G 2E7, Canada

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. To allow for extended viewing time, posters will remain on display until 5:00 p.m.

Contributed Papers

2aSC1. Information theory and variance estimation techniques in the analysis of category rating and paired comparisons. Sumiko Takayanagi (Dept. of Linguist., UCLA, Los Angeles, CA 90095-1543), Mark Hasegawa-Johnson (MIT, Cambridge, MA 02139), Laurie S. Eisenberg, and Amy Schaefer Martinez (House Ear Inst., Los Angeles, CA 90057)

Speech clarity judgments of bandpass filtered sentences were measured by category rating and paired comparisons in five age groups—4, 5, 6, 7, and 8 years old, and adults [data for children were provided from Eisenberg and Dirks, J. Speech Hearing Res. **38**, 1157–1167 (1995)]. The judgment data within each age group were subjected to information theory and variance estimation analytical techniques in order to investigate developmental differences in task performance capability. All response variability was converted into equivalent signal detection ''error'' components, representing the ability of subjects in a given age group to perform a given psychophysical procedure. Preliminary results revealed a different developmental trend for each psychophysical procedure. For category rating, children 4 to 6 years of age exhibited higher error components in their responses, suggesting reduced capability to extract information compared with adults. In contrast, children as young as 5 years of age were approaching adultlike low error components in the paired comparisons. In this particular study, paired comparisons may provide a greater amount of information with less variance than category rating when subjective judgments of sound quality are measured. [Work supported by NIH.]

2aSC2. Must sine wave analogs be perceived as implicit speech to be perceived categorically? John Kingston and Cecilia J. Kirk (Linguist. Dept., Univ. of Massachusetts, Amherst, MA 01003)

Sawusch and Gagnon [J. Exp. Psychol., HPP **21**, 635–652 (1995)] obtained categorical responses to /la/–/ra/ sine wave analogs when listeners who heard the stimuli as nonspeech were frequently reminded of $F3$'s range or extremes. This result contrasts with responses to sine wave /la/–/ra/ obtained by Best *et al.* [Percept. Psychophys. **45**, 237–250 (1989)] and responses to sine wave /wa/–/ja/ obtained by Johnson and Ralston [Phonetica **51**, 195–209 (1994)]. Only those listeners who heard the analogs as (distorted) speech responded categorically. Instead of comparing speech and nonspeech listeners, the experiments tried to create an implicit expectation that sine wave analogs are (distorted) speech by playing listeners speech first. Listeners identified and discriminated members of sine wave /la/–/ra/ after identifying and discriminating either members of the original /la/–/ra/ speech continuum or of the same sine wave continuum. Other groups of listeners heard speech or sine wave /wa/–/ja/ before the sine wave /la/–/ra/; their performance allows us to separate acoustic from type similarity. If after hearing speech, listeners hear the analogs as implicit speech, they should perceive them more categorically than listeners who hear analogs first. Results show instead that listeners responses are as strongly categorical to the sine wave /la/–/ra/ in the second phase after all preceding stimulus types, even though the analogs are not heard as speech. Categorical perception thus does not depend on an implicit speech percept, nor are frequent reminders required to elicit it to sine wave analogs. [Work supported by NIH.]

2aSC3. Analysis of formant transition and burst cue integration in stop consonant perception. Krishna K. Govindarajan (Speech Commun. Group, RLE, MIT, Cambridge, MA 02139)

While formant transitions and burst cues have been implicated in the place perception of stop consonants, the exact nature of the integration of these cues remains unknown. This study investigates the interaction of these cues by having listeners identify synthetic consonant-vowel (CV) stimuli in which the second formant ($F2$) transition and the burst center frequency are simultaneously varied in both a front and back vowel context. The resulting identification surface will be presented for the individual subjects, and listeners' categorizations will be compared to categorization using discrimant analysis and a neural model. Preliminary evidence suggests that the burst center frequency is less critical to identification than which formant it is affiliated with, e.g., $F2$ or $F3$. Moreover, in certain contexts, the formant affiliation of the burst overrides the $F2$ transition information. This result seems to suggests that listeners are interpreting the burst as a formant that starts prior to the other formants, i.e., a ''leading'' formant. To study this further, the results from another experiment employing synthetic CV stimuli, where one formant leads the other formants, will be presented. [Work supported by NIH Grant DC00194.]

2aSC4. Influence of phonotactic rules on perception of ambiguous segments. Elliott Moreton (Dept. of Linguist., Univ. of Massachusetts, Amherst, MA 01003)

In every language, some sequences of sounds are illegal. English, for instance, bans stressed lax vowels word-finally—[bI] cannot be an English word. Linguists traditionally attribute this to language-particular phonotactic rules. More recently, some psychologists have suggested instead that phonotactics is the emergent statistical property of the lexicon, caused by the frequency of some sequences and the rarity of others. The issue is tested here by comparing the effects of absolute phonotactic rules versus nonphonotactic lexical frequency differences on phonetic category boundaries. Stimuli are disyllabic English pseudowords ending in a stressed syllable whose vowel is ambiguous. One continuum is [gri] (very common in that position) to [grI] (illegal); another is [kri] (legal but very rare) to

[krI] (illegal). Controls, with both end points legal, are [grich]-[grIch] and [krich-krIch]. The [I] end point's illegality should move the i–I boundary towards [I] compared with the controls. The rule theory predicts equal shifts for the [gr] and [kr] continuua; the statistical theory says it will be much larger for the frequent [gr]. [Work supported by NIH.]

2aSC5. Phonetic variability in /s/ and /z/ segment errors: A multitalker acoustic study. Richard Wright and Stefan Frisch (Speech Res. Lab., Dept. of Psych., Indiana Univ., Bloomington, IN 47405)

Phonological speech errors are a traditional source of evidence for phonological units, such as features and segments, in both linguistic and psycholinguistic theory. Assumptions about the unitary nature of speech errors have been brought into question by research that found evidence for subfeatural errors in speech production [R. Mowrey and I. MacKay, J. Acoust. Soc. Am. **88**, 1299–1312 (1990)]. This poster presents data from an ongoing acoustic study of speech errors in /s/ and /z/ productions. Previously reported findings indicate that the errors involve either single or multiple changes in periodicity, amplitude and duration of frication, and vowel rise time [S. Frisch and R. Wright, J. Acoust. Soc. Am. **101**, 3178(A) (1997)]. The data are tape recordings of an error-inducing experiment using nonsense tongue twisters. Recordings of six talkers producing four different tongue twisters targeting /s/ and /z/, for examples, ''sit zap zoo sip,'' were digitized and analyzed. This study examines the phonetic variation in featural and subfeatural errors across talkers and segmental contexts. Results suggest that details of individual talker's production strategies and the phonetic context influence the distribution of featural and subfeatural errors. [Work supported by NIH Grant No. DC00012.]

2aSC6. Acoustic analysis of simple vowels with nasal coda in standard Chinese. Marilyn Y. Chen (Inst. of Linguist., Chinese Acad. of Social Sci., 5 Jianguomennei Dajie, Beijing, China 100732, linmc@sun.ihep.ac.cn)

More than 40% of the monosyllables in standard Chinese (SC) have a vowel followed by a nasal coda. The goal of this study is to determine by acoustic analysis how the vocalic portion is affected by the nasal context due to coarticulation in SC disyllabic words. Simple vowels /i–a–e/ in two types of syllables were examined: (1) with a nasal coda /n–ng/ followed by a syllable beginning with a stop or vowel, and (2) without a nasal coda followed by a syllable beginning with a stop or /n/. Spectral measurements were made every 10 ms in the vowel and the nasal consonant. The maximum first difference for the first four formant amplitudes was used to detect the vowel–nasal consonant boundary. The formant trajectories in the vowel over time were used as indicators for the place of the nasal coda formed with or without oral closure. Contours of adjusted $A1-P0$ and $A1-P1$ (amplitude differences between the first formant and nasal peaks) over time showed that vowels with nasal coda are more nasalized than vowels in a non-nasal context or in an intersyllabic nasal context. [Work supported by MIT International Science and Technology Initiative.]

2aSC7. Exploration of the perceptual magnet effect using the mismatch negativity auditory evoked potential. Anu Sharma and Michael F. Dorman (Arizona State Univ., Dept. of Speech and Hearing Sci., Tempe, AZ 85287-0102, anu.sharma@asu.edu)

This study assessed the replicability of the ''perceptual magnet effect (PME)'' [Kuhl, Percept. Psychophys. **50**, 93–107 (1991)] and investigated underlying neurophysiologic processes using the mismatch negativity (MMN) evoked potential. A stimulus continuum from /i/–/e/ was synthesized by varying $F1$ and $F2$ in equal mel steps. Subjects categorized, rated the goodness of stimuli, and discriminated between stimulus pairs. Results revealed that Kuhl's prototype was identified and rated as a good (albeit, not the best) exemplar of /i/, whereas, Kuhl's nonprototype (NP) was identified as /i/ only 20% of the time, and therefore, was not a suitable NP. In this study, the prototype was the stimulus with the lowest $F1$ and highest $F2$ values and the NP was close to the category boundary. Discrimination accuracy was not significantly different in the prototype and NP conditions. That is, no perceptual magnet effect was observed. Furthermore, despite equal mel differences between the stimulus pairs, the MMN was largest for the pair with lowest $F1$ and highest $F2$ values. Therefore, the MMN appears to be sensitive to within category acoustic differences. The behavioral and electrophysiologic results indicate that discrimination of stimuli near a prototype is based on the auditory structure of those stimuli.

2aSC8. The effect of presentation level on the perception of voicing. Cecilia J. Kirk (Linguist. Dept., Univ. of Massachusetts, Amherst, MA 01003)

Kluender *et al.* [J. Acoust. Soc. Am. **97**, 2552–2567 (1995)] propose that voicelessness in a continuum from short to long lag VOTs is signaled by abrupt synchrony capture of $F2–F3$ tuned fibers by $F1$ when voicing begins. Their hypothesis also predicts more capture at higher presentation levels, producing crossover to voiceless judgments at shorter VOTs, an effect observed in their experiments. In an attempt to replicate Kluender *et al.*'s findings, listeners were presented with a /ba/–/pa/ continuum at 48, 60, and 72-dB SPL: VOT crossovers did not differ significantly between the two highest levels, but were significantly longer for the lowest compared to the two highest levels. A second experiment tested manipulations of relative aspiration level by combining aspiration levels of 42, 48, 54, 60, and 66 dB with a 72-dB vowel; VOT crossovers increased with decreasing aspiration level. The results of these two experiments are compatible with the alternative hypothesis that reducing aspiration level below a critical level makes aspiration effectively inaudible. In any case, the synchrony capture hypothesis fails on two counts: Sinex and MacDonald [J. Acoust. Soc. Am. **85**, 1995–2004 (1989)] observe that capture of mid- and high-CF neurons by $F1$ does not vary with VOT for VOTs less than 50–60 ms. In addition, Lotto and Kluender [J. Acoust. Soc. Am. **99**, 2591(A) (1996)] demonstrate that the level effect still obtains when $F1$ is presented to a different ear than the higher formants.

2aSC9. Integration of acoustic information in the perception of [s] ± stop clusters by children and adults. Court S. Crowther and Susan Nittrouer (Boys Town Natl. Res. Hospital, 555 North 30th St., Omaha, NE 68131)

To compare the integration of acoustic information in phonetic perception by young children and adults, second formant frequency at voicing onset ($F2$ onset) and gap duration were varied as cues distinguishing ''spa'' and ''sa.'' Listeners performed same–different (AX) discrimination in four conditions. The gap of the standard (A) stimulus was always consistent with ''spa;'' gaps of the comparison (X) stimuli decreased in duration along a continuum, thus becoming more consistent with ''sa.'' In the two ''one-cue'' conditions, $F2$ onset was the same for the A and X stimuli. In the ''two-cue cooperating'' condition, $F2$ onset for stimulus A supported ''spa,'' and $F2$ onset for X stimuli supported ''sa.'' In the ''two-cue conflicting'' condition, $F2$ onset for stimulus A supported ''sa,'' and $F2$ onset for X stimuli supported ''spa.'' Results led to three conclusions: (1) Child and adult discrimination performance was similar overall, suggesting that the child-adult differences across conditions were not due to differences in general auditory ability. (2) The acoustic cues were integrated in phonetic perception, but acoustic differences were nonetheless accessible perceptually for AX discrimination. (3) However,

these acoustic differences are less accessible to adults than to children after integration in phonetic perception. [Work supported by research Grant No. 5 RO1 DC 00633 from the NIDCD/NIH.]

2aSC10. Perceptual-magnet effect as categorical perception. Andrew J. Lotto (Dept. of Psych. and Parmly Hearing Inst., Loyola Univ. Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626, ajlotto@facstaff.wisc.edu), Lori L. Holt, and Keith R. Kluender (Univ. of Wisconsin, Madison, WI 53706)

The perceptual-magnet effect has received a great deal of attention in recent years as a possible foundation for phonetic category acquisition. This effect is based upon studies showing that category goodness affects intracategory discriminability. Unfortunately, there is a possible confound in many of these demonstrations. Whereas category identity and goodness ratings are collected for stimuli presented in isolation, discrimination scores are obtained for stimulus pairs. In the present study, vowel identification and goodness ratings were obtained from the same stimulus pairs that would later be presented in an AX discrimination task. Stimuli were identical to the vowel vector used by Iverson and Kuhl [J. Acoust. Soc. Am. **97**, 553–561 (1995)]. Identification functions varied substantially as a function of the pairings in a manner that may be described as contrastive. Vowels labeled as /e/ when presented with a good /i/ exemplar were labeled as /i/ when presented with a good /e/. These identification functions were used to predict discrimination scores based solely on perceived phonemic identity. Predictions accounted for all of the difference in discriminability between prototypical and nonprototypical exemplars. Perceptual-magnet effects appear to be little more than classic categorical perception. [Work supported by NSF.]

2aSC11. Effects of the uvula and the epiglottis on measurements of vowel production: Assessment by synthesis. Harriet Magen (Haskins Labs., 270 Crown St., New Haven, CT 06511 and Dept. of Commun., Rhode Island College, Providence, RI 02908) and Min Kang (Haskins Labs., New Haven, CT 06511)

The acoustic shape of vowels is determined by the cross-sectional areas along the vocal tract. These areas can be predicted from the midsagittal distance, given certain assumptions about the relation of that distance to the area [P. Mermelstein, J. Acoust. Soc. Am. **53**, 1070–1082 (1973)]. The present study used midsagittal MRI images of the vocal tract during sustained vowel production to provide areas as input to an articulatory synthesizer. The acoustic consequences of two methods of measuring distance will be assessed: computing distance with and without the epiglottis and with and without the uvula. For some vowels, the uvula extended into the oral cavity to different degrees across repetitions, resulting in large within-vowel variability for that region. Similarly, distances in the region of the epiglottis depended on the degree to which the epiglottis is detached from the root of the tongue, though this tended to be more consistent across repetitions. Closer matches, both in terms of formant frequencies and perceptual similarity, will indicate which method best represents the effective area predicted by the midsagittal distance. [Work supported by NIH Grant DC-02717.]

2aSC12. Endoscopic and acoustic evidence of the guttural natural class. Bushra Zawaydeh (Indiana Univ., Dept. of Linguist., Memorial Hall 322, Bloomington, IN 47403)

While Arabic gutturals comprise of laryngeals, pharyngeals, uvulars, and emphatics, Salish gutturals do not include laryngeals. To see if the pharynx could be the articulator that unites gutturals, an endoscopic experiment was conducted on a native speaker of Arabic. The pharyneal movement was videotaped while the subject repeated nonwords that had a plain or a guttural sound in the context of ?aCa. The result is that there is a significant difference in the pharyngeal width between the plain sounds and the pharyngeal, emphatic, and uvular sounds. As for the laryngeals, they showed no constriction in the pharynx. This result could explain why Salish laryngeals do not pattern with other gutturals, but it does not explain why Arabic laryngeals pattern with the rest of the gutturals. Hence, an acoustic analysis of the voice of four native speakers of Arabic saying the same nonwords that were used above revealed that the midpoint steady state of the second low vowel has a higher $F1$ in all of the guttural sounds, including the laryngeals, than the $F1$ of the plain sounds. This means that in Arabic (unlike Salish), the guttural natural class is better explained acoustically than articulatorily.

2aSC13. Effects of consonant environment on vowel formant patterns. James M. Hillenbrand and Michael J. Clark (Speech Pathol. and Audiol., Western Michigan Univ., Kalamazoo, MI 49008)

A significant body of evidence has accumulated indicating that vowel identification is influenced by spectral change patterns. For example, a large scale study of vowel formant patterns showed substantial improvements in category separability when a pattern classifier was trained on multiple samples of the formant pattern rather than a single sample at steady state [Hillenbrand *et al.*, J. Acoust. Soc. Am. **97**, 3099–3111 (1995)]. However, in the earlier study all utterances were recorded in a constant /hVd/ environment. The purpose of the present study was to determine whether a close relationship between vowel identity and spectral change patterns is maintained when the consonant environment is allowed to vary. Recordings were made of six men and six women producing the vowels /i,ɪ,ɛ,æ,ɑ,ʌ,ʊ,u/ in isolation and in CVC syllables. The CVC utterances consisted of all combinations of seven initial consonants (/h,b,d,g,p,t,k/) and six final consonants (/b,d,g,p,t,k/). Formant frequencies for F_1–F_3 were measured every 5 ms during the vowel nucleus using an interactive editing tool. Results showed highly significant effects of phonetic environment. As with an earlier study of this type, particularly large shifts in formant patterns were seen for rounded vowels in alveolar environments [K. Stevens and A. House, J. Speech Hear. Res. **6**, 111–128 (1963)]. Despite these context effects, large improvements in category separability were observed when a pattern classifier incorporated spectral change information. [Work supported by NIH.]

2aSC14. Acoustic differences between syllable-initial [f] and [θ]. Marnie E. Miller and Susan Nittrouer (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131)

The consonants [f] and [θ] are notoriously confusable in perception. Consistent with this fact, previous acoustic analyses have failed to demonstrate strong differences between syllables containing the two consonants in syllable-initial position. However, these earlier studies were limited in the numbers of speakers providing samples and in the measurements made. Therefore, five tokens of each of the syllables [fɑ], [fu], [θɑ], and [θu] were recorded by each of eight speakers (four men and four women) and five acoustic properties measured in each syllable in the vicinity of the fricative segment: duration, amplitude, and first spectral moment of the fricative noise, and the frequency of the second and third formants at voicing onset. No consistent differences were observed for the noises themselves. Differences between syllables resided in the second and third formant transitions at voicing onset: frequencies of both formants are lower for [f] vowel than for [θ] vowel, both for males and females and for [ɑ] and [u]. Thus acoustic evidence is provided supporting the perceptual result that the voiced formants provide whatever acoustic information is available for making phonetic judgments of these consonants. [Work supported by research Grant No. 5 R01 DC 00633 from the NIDCD/NIH.]

2aSC15. Word–initial versus word–final consonant articulation. Patricia A. Keating (Phonet. Lab., Linguist. Dept., UCLA, Los Angeles, CA 90095-1543, keating@humnet.ucla.edu)

The position of a consonant within its prosodic word is known to have an effect on the degree, or strength, of its articulatory stricture: in stronger prosodic positions, the active articulator may be higher, making the oral opening smaller. This study examines the robustness of this asymmetry in consonant articulation. Electropalatography was used with five speakers to determine linguo–palatal contact for several English consonants in word–initial versus word–final position. The various experiments manipulated such factors as identity of the consonant, the test word's lexical stress, its position in its sentence, and phrasal accent in the sentence. Results show that: (1) noncontinuant consonants generally have more contact in word–initial than word–final position, but sibilant fricatives do not; (2) intonational–phrase–final position is a strong position, so that word–final consonants /t d n/, which are also sentence final often show little word–final weakening (/l/ is an exception for some speakers); and (3) in contrast to this positional effect, lexical stress and phrasal accent never strengthen word–final consonants in this way, nor do these prosodic factors consistently strengthen word–initial consonants. [Work supported by NSF.]

2aSC16. Categorical perception of Japanese moraic phonemes. Teruhisa Uchida (Dept. of Linguist., Ohio State Univ., Columbus, OH and Res. Div., Natl. Ctr. for Univ. Entrance Exam., Tokyo 153-8501, Japan, uchida@rd.dnc.ac.jp)

This study found that segment duration is perceived categorically in Japanese. Using a speech time-scale modification algorithm, PICOLA plus2, the nonword/kesonato/pronounced by a native Japanese male was manipulated to generate stimulus continua. The durations of /s/, /n/, /a/, and /t/ closure interval were manipulated in four continua. In the first experiment, three stimulus pairs, which had the same disparity in duration were selected from the short, long, and middle region of each continuum. Forty-eight randomized trials were presented to 20 native Japanese, who were asked individually to distinguish the difference of sounds by the ABX method. Pairs from the short and long regions of the continua were not accurately discriminated while pairs from the middle region were. In the second experiment, 44 stimuli were selected from every continuum. They were presented to the subjects for identification in Hiragana (Japanese characters). The second result demonstrated the existence of a moraic phoneme boundary paralleling the discrimination peak in the first experiment. This research concluded that native Japanese perceive the duration of relatively steady-static sounds categorically. And also, it is suggested that native Japanese utilize an automatic duration detection routine.

2aSC17. The effect of speaking-rate manipulations on segment durations and formant frequencies in dysarthric speakers with amyotrophic lateral sclerosis. Jacqueline S. Laures and Jing-Yi Jeng (Dept. Commun. Disord. and Waisman Ctr., 1500 Highland Ave., Univ. of Wisconsin, Madison, WI 53705, jslaures@students.wisc.edu)

The effect of speaking-rate manipulations on dysarthric speech is interesting from both theoretical and clinical perspectives. The purpose of the present study was to examine the influence of speaking-rate adjustments (conversational versus fast) on segment durations and vowel formant frequencies derived from sentence productions in speakers with amyotrophic lateral sclerosis (ALS), a disease that involves death of motor neurons serving the speech mechanism, with subsequent loss of muscle mass and strength. Speakers with ALS were clearly able to change their rate for the fast condition, though not in the same proportion as the control speakers. The segment durations responded to the rate manipulation in much the same way. For formant frequencies, the effect of rate was complex, but may be generally summarized as follows: $F1$ was affected systematically by the rate manipulation in control males, but not among the male subjects with ALS; for females in both groups, there was little effect of rate on $F1$. The effects of rate on $F2$ measures showed no clear trend, varying across vowels and groups. Results will be discussed relative to the speech production deficit in ALS and acoustic models of speech intelligibility. [Work supported by NIDCD Award DC00319.]

2aSC18. Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. Jing-Yi Jeng, Jacqueline S. Laures, and Gary Weismer (Dept. Commun. Disord. and Waisman Ctr., Univ. of Wisconsin, Madison, WI 53705-2280)

Although sentence production is more common than single-word production in daily life, acoustic and intelligibility data from persons with neurogenic speech disorders is largely limited to single-word and target-word utterances. The purpose of the current study is to examine sentence productions of individuals with Parkinson's disease (PD) and amyotrophic lateral sclerosis (ALS), and to compare them to data obtained from neurologically normal individuals. Each speaker produced multiple repetitions of a small set of sentences. Acoustic measures included overall sentence durations, segment durations, and vowel formant frequencies. Speech intelligibility of the sentences was measured using scaling techniques. For some subjects, previously obtained acoustic and intelligibility data for single words will be related to the findings for sentences. Results will be discussed descriptively as deviations from patterns observed in normal individuals, and in terms of associations between the acoustic and intelligibility data. [Work supported by NIDCD Award DC00319.]

2aSC19. Articulatory trade-offs reduce acoustic variability in American English /r/ productions. Frank H. Guenther (Dept. of Cognit. and Neural Systems, Boston Univ., Boston, MA 02215), Carol Y. Espy-Wilson, Suzanne E. Boyce, Melanie L. Matthies (Boston Univ., Boston, MA 02215), Majid Zandipour, and Joseph S. Perkell (MIT, Cambridge, MA 02139)

Acoustic and articulatory recordings indicate that speakers utilize trade-offs between different articulatory parameters to maintain acoustic stability during American English /r/ productions. Electromagnetic midsagittal articulometer data from seven subjects producing /r/ in five phonetic contexts were analyzed. Articulator configurations for /r/ in different contexts showed systematic trade-offs between constriction length, constriction area, and front cavity volume. The hypothesized effects of these three parameters on $F3$, derived from acoustic theory, were verified using a 51-tube model of the vocal tract derived from structural MRI scans of a speaker producing /r/. All seven subjects utilized a trade-off between front cavity volume and the length of the constriction formed by the tongue. Further analysis of acoustic and articulatory variabilities indicates that the net effect of these articulatory covariances is a considerable reduction in acoustic variability, thus allowing large contextual variations in vocal tract shape; these contextual variations in turn appear to reduce the total amount of articulatory movement required. These findings contrast with the view that speaking involves a target set of vocal tract constrictions for each phoneme and provide strong evidence for acoustic theories of speech production. [Work supported by NIDCD, NSF, and the Sloan Foundation.]

2aSC20. Spectral transition in the perception of English segments. Natasha Warner (Dept. of Linguist., Univ. of California Berkeley, 2337 Dwinelle Hall, Berkeley, CA 94720, nwarner@uclink.berkeley.edu)

There is a long-standing debate on whether speech is perceived through steady states or transitions [D. Kewley-Port and D. Pisoni, J. Acoust. Soc. Am. **73**, 1779–1793 (1983)]. Furui [S. Furui, J. Acoust. Soc. Am. **80**, 1016–1025 (1986)], gating Japanese monomoraic syllables, found listeners identified syllables correctly when they heard the point of maximal spectral transition (maximum D, a measure based on cepstral coefficients). In this paper, this work is extended to English, and to a wider variety of transitions. 128 English words, representing transitions between all manners of articulation, were gated at 20-ms intervals. VC, CC, and VV transitions were used in addition to Furui's CV and CjV syllables. The

results show that for almost all transitions (diphones), there is a clear point at which listener's identification of the second segment becomes highly accurate. For many diphones (54%), this point is within 15 ms of the maximum D. However, for other diphones (especially those with far-reaching perceptual cues), the point of recognition is farther from the maximum D. This paper will discuss reasons for the discrepancies between Furui's experiment and the current one, including language-specific factors, and will address implications for the dynamic theory of speech perception.

2aSC21. Spectral change in the vowel formant space of Long Island vowels. Marie K. Huffman and Elyse E. Tamberino (Dept. of Linguist., SBS S201, SUNY, Stony Brook, NY 11794-4376, mhuffman@ccmail.sunysb.edu)

Hillenbrand *et al.* [J. Acoust. Soc. Am. **97**, 3099–3111 (1995)] have shown that English vowels are distinguished better by discriminant analysis when a measure of spectral change is used than when a simple formant steady-state measure is used. Many English speakers on Long Island have six front vowels instead of the five treated in most discussions of American English—in addition to low front æ, there is a higher, or ''tensed'' æ (henceforth AE). With this more crowded vowel space, spectral dynamism should be equally, if not more, important to distinguishing Long Island vowels than other varieties of English such as the Michigan English described by Hillenbrand *et al.* Vowels were recorded in real words of the form hVd and bVd, spoken in a frame sentence. $F1$ and $F2$ values from LPC spectral analysis were analyzed via formant plots and discriminant analysis. Preliminary results for front vowels support the hypothesis: improvement in discrimination using dynamic rather than static spectral measures was largest for ϵ and AE, the vowels that are closest in the vowel space. Overall use of the vowel space, including comparisons with other dialects and studies will also be presented. [Work supported by NSF.]

2aSC22. Cross-validation study of vowel perception in Finnish and English. Terrance M. Nearey and Michael Kiefte (Dept. of Linguist., Univ. of Alberta, Edmonton, AB T6G 2E7, Canada)

A large stimulus continuum (972 stimuli in an $F1$–$F2$–$F3$ space) was categorized by both Finnish and English listeners. Further results of logistic regression modeling of listeners' responses using formant frequencies and/or a spectral shape representation (warped−frequency LPC cepstral coefficients) are reported. Cross-validation model selection methods [J. Shao, J. Am. Stat. Assoc. **88**, 486−494 (1993)] indicate that a combination of the two parameter sets (formant and cepstral) is superior to either alone. However, for both languages the best prediction scores were obtained from a purely formant-based analysis that included squares and cross products of frequency values, thereby allowing for quadratic category boundaries in the formant space. Analysis to date provides no evidence that spectral-shape information (versus formant frequencies alone) plays a role in listeners' perception of these stimuli in either Finnish or English. Results of additional analyses will be presented at the meeting. [Work supported by SSHRC.]

2aSC23. Perceptual interactions among voice onset time, second formant onset, voice, and place of articulation. José R. Benkí (Program in Linguist., Univ. of Michigan, 1076 Frieze Bldg., Ann Arbor, MI 48109-1285, benki@umich.edu)

The categorization of voicing in syllable-initial stops is dependent primarily on voice onset time (VOT), while the categorization of place of articulation (POA) in stops is dependent primarily on higher formant transitions. This study tested for perceptual interactions among these two orthogonal acoustic dimensions (VOT and higher-formant transitions) and the corresponding phonological contrasts of voice and POA. Fifteen subjects categorized initial stops in a /Ca/ context as /ba/, /pa/, /ga/, or /ka/. The initial stop of each token was from a synthetic two-dimensional continuum with acoustic dimensions of VOT (0–45 ms) and second formant ($F2$) onset frequency (1100–1700 Hz), so that the corners of the continuum were unambiguous /ba/, /pa/, /ga/, and /ka/. A model of POA and voicing categorization derived from logistic regression suggests that the VOT category boundary can change depending on the perceived POA. Tokens categorized as bilabial were more likely than those perceived as velars to be also categorized as voiced, regardless of the $F2$ onset. The change in perceptual VOT boundary is consistent with previous results [J. R. Benki, J. Acoust. Soc. Am. **100**, 2691 (1996)], though the magnitude of the effect is not as large. [Work supported by NSF.]

2aSC24. Effect of relative amplitude manipulation on perception of voiceless fricatives by normal and impaired listeners. Mark S. Hedrick (Dept. of Audiol. and Speech Pathol., Univ. of Tennessee, Knoxville, TN 37996)

In the present study, previous work on the relative amplitude and presentation level was extended to include the voiceless fricative /s/–/ʃ/ contrast. Synthetic CV stimuli were used, in which format transitions, frication duration, and relative amplitude were manipulated. For the relative amplitude manipulation, the amplitude of friction relative to vowel onset in the $F3$ frequency range was varied across a 20-dB range. Frication duration was varied from 140–50 ms. Stimuli were presented at 90 dB SPL. Eight listeners with normal hearing and three listeners with moderate sensorineural hearing loss participated in the experiment. Results showed that listeners with sensorineural hearing loss gave significantly more /s/ responses for lower relative amplitude values. Results from this study and previous studies [Hedrick *et al.*, J. Acoust. Soc. Am. **98**, 1292–1303 (1995); M. S. Hedrick and W. Jesteadt, J. Acoust. Soc. Am. **100**, 3398–3407 (1996)] suggest that perception of acoustic cues for consonants by listeners with sensorineural hearing loss may differ from that of listeners with normal hearing, even when the cues are audible.

2aSC25. Perception of long and short tones in Taiwanese speech. Francine H. Jian (Dept. of Linguist., Univ. of Reading, Reading RG6 6AY, England)

Taiwanese, in common with other ancient Chinese dialects, has a more complex tonal structure than the well-known standard Peking Mandarin. One important feature is the notion of ''long'' and ''short'' tones. In order to build a successful computer-aided speech recognition system for Taiwanese speech, it is vital to understand this much overlooked phenomena. In a set of experiments, we set out to determine how subjects distinguish long versus short tones. A palette of long and short words was composed such that a long word would have a short counterpart and vice versa. Using a resynthesis tool, these words were gradually modified to resemble the opposite tone to study the boundaries of perception. Two main features were investigated, namely word duration and energy drop-off (or glottal stop). The experiments reveal that duration has an insignificant impact on the distinction between long and short tones, while the energy drop-off has a strong impact on the perception of long versus short tones. A word with a steep energy drop-off is more likely to be perceived as a short tone than a word with a gentle energy drop-off.

2aSC26. Factors affecting vowel identification in hearing-impaired listeners. Amy T. Neel (Dept. of Speech and Hearing Sci., Indiana Univ., Bloomington, IN 47405)

Vowel identification generally poses little difficulty for listeners with mild to moderate hearing impairment despite evidence for poorer frequency resolution compared to normal-hearing listeners. Normal-hearing listeners use formant movement and duration cues as well as spectral target cues to identify vowels. This study examines use of formant movement and duration cues by hearing-impaired listeners. Sets of very natural-sounding vowel stimuli with and without formant movement cues (dy-

namic and static) and with and without duration cues (appropriate and fixed) were made with a new speech resynthesis technique [H. Kawahara, Proc. ICASSP, 1–4 (1997)] using vowels produced by male and female talkers in /dVd/ context. Two groups of young normal-hearing listeners, one with simulated hearing loss typical of 70 year-old males and one control group, displayed no difference in overall vowel identification scores. Both groups obtained significant benefit from tokens containing formant movement cues. Duration was not as effective a cue for vowel identification as formant movement. Results for elderly hearing-impaired and elderly normal-hearing listeners will be presented, and the impact of talker variability upon vowel identification will be discussed. [Work supported by NIH-NIDCD 02229.]

TUESDAY MORNING, 2 DECEMBER 1997 FRIARS/PADRE ROOMS, 8:00 A.M. TO 12:20 P.M.

Session 2aSP

Signal Processing in Acoustics, Psychological and Physiological Acoustics and Speech Communication: Acoustics in Multimedia I: Perceptual Issues

David I. Havelock, Chair
Institute for Microstructural Sciences, National Research Council, M-36 Montreal Road, Ottawa, Ontario K1A 0R6, Canada

Chair's Introduction—8:00

Invited Papers

8:05

2aSP1. Virtual reality: The perils of synthesis that is almost correct. William M. Hartmann (Phys. Astron., Michigan State Univ., East Lansing, MI 48824)

The experimental method of Hartmann and Wittenberg [J. Acoust. Soc. Am. **99**, 3678–3688 (1996)] is an approach to virtual reality with many limitations, but it is the most reliable technique for scientific purposes. Experiments using this method find that a precise VR synthesis is entirely convincing. Of interest are the perceptual consequences of perturbations on the precise synthesis. Large, random perturbations lead to diffuse and internalized images. Better controlled perturbations can lead to split images with perceived timbre differences and even pitch differences. Incorrect images overwhelmingly tend to be localized in the back, e.g., reversing interaural time differences while maintaining interaural intensity differences reverses the azimuth but also reflects the image in a front–back plane (when not internalized). Smaller perturbations produce images that are unitary, but not compact, and they are normally located at distances that are nearer than the true source. Perturbations of intermediate size can lead to extended but well-defined source images, such as gobs of sound extending from the back of the neck and running down the spine. Spatial effects like these can be disconcerting and illustrate the perils of synthesis, getting it almost right, but not entirely right. [Work supported by the NIDCD, DC00181.]

8:35

2aSP2. Perceptual validation of multimedia sound systems. Fred Wightman and Doris Kistler (Waisman Ctr., Univ. of Wisconsin, Madison, WI 53705, wightman@waisman.wisc.edu)

Claims about the subjective performance of consumer audio products are rarely supported by psychophysical data. Advertising hype and anecdotal reports from product reviewers are typically used in place of scientific evidence. With 3-D audio products, however, research preceded product development, and existing psychophysical data address many issues related to potential product performance. These data suggest that integrating high-fidelity 3-D sound into multimedia systems is likely to be a formidable challenge. One problem is that the acoustics of the listening environment are intractable unless the listener is constrained in some way (e.g., forced to wear headphones). Another obstacle is that even if the acoustics are handled properly, large individual differences preclude generalizations about the costs/benefits of engineering compromises, such as the use of nonindividualized spatial processing. Perhaps most important, the data suggest that the perceptual errors likely to be produced in any less-than-perfect 3-D audio system are not subtle; sound images fail to externalize or are heard in the rear when they are intended for the front. These problems underscore the importance of using rigorous psychophysics to evaluate the perceptual adequacy of any multimedia sound system. [Work supported by NASA, ONR, NIH, and Rockwell Semiconductor Systems.]

9:05

2aSP3. Perceptual fidelity versus engineering compromises in virtual acoustic displays. Elizabeth M. Wenzel (NASA-Ames Res. Ctr., M.S. 262-2, Moffett Field, CA 94035-1000, bwenzel@mail.arc.nasa.gov)

Immersive, multisensory displays are increasingly a goal of advanced human–machine interfaces, and three-dimensional sound techniques are now practical enough to be applied to these displays. The goal of virtual acoustics is to simulate the complex acoustic field experienced by a listener freely moving within an environment. Of course, such complexity, freedom of movement, and interactivity is not always possible in a ''true'' virtual environment, much less in lower-fidelity multimedia systems. Many of the perceptual and engineering constraints (and frustrations) that researchers, engineers, and listeners have experienced in virtual audio are relevant to multimedia. In fact, some of the problems that have been studied will be even more of an issue for lower fidelity systems

that address the requirements of a large, diverse, and ultimately unknown audience. Examples include individual differences in head-related transfer functions, a lack of real interactivity in many multimedia systems, and perceptual degradation due to low-sampling rates or compression. This paper discusses some of the engineering constraints and perceptual consequences faced during implementation of virtual acoustic displays for NASA applications such as aeronautics and shuttle launch communications. An attempt is also made to relate these issues to low-fidelity applications such as the Internet.

9:35

2aSP4. Sound source determination and segregation. William A. Yost (Parmly Hearing Inst., Loyola Univ., 6525 N. Sheridan Rd., Chicago, IL 60201)

In any medium the auditory system allows one to determine objects in our environment based on the sounds they produce. In a multisource acoustic environment the sounds produced by the sources arrive at the ears as a single sound field. The neural and perceptual processes of the auditory system parse this complex sound field into perceptions of the individual sound sources. There are several key properties of sound that allow the auditory system to determine the aspects of the complex sound field that are associated with each contributing sound source. In order for individual sound sources to be perceived in any medium, these properties must be well preserved in the sound reproduction. This talk will describe the parameters responsible for sound source determination and segregation and one's ability to perceive individual sound sources in a complex, multisource environment. [Work supported by NICDC, DC000290.]

10:05–10:20 Break

10:20

2aSP5. Adapting to discrepant information in multimedia displays. Barbara Shinn-Cunningham (Dept. of Cognit. and Neural Systems, Boston Univ., Rm. 311, 677 Beacon St., Boston, MA 02215)

The ultimate goal of multimedia displays is to generate a realistic and engaging display of visual, auditory, and/or proprioceptive information that creates a consistent view of the ''rendered'' world. Unfortunately, time delays and inaccuracies in the display system may cause events to be temporally or spatially misaligned across modalities. Even within a single modality, technological constraints may limit the accuracy of the display. If a system is for entertainment, such discrepancies will matter only when they are consciously perceived by the observer. If the display is for command and control applications (for instance, to convey information to an air traffic controller or to the operator of a remotely controlled vehicle), discrepancies matter whenever they decrease performance on the task of interest, especially if such decrements in performance cannot be overcome by training. A series of experiments investigating the degree to which subjects can adapt to discrepancies between normal experience and simulated auditory spatial information suggests that subjects can adapt to such discrepancies; however, the degree to which adaptation affects the ability of the subject to perform spatial tasks depends on the amount and type of spatial discrepancy present in the display. [This work supported by AFOSR.]

10:50

2aSP6. Acoustics and psychoacoustics of desk-top sound systems. Tomlinson Holman and Chris Kyriakakis (TMH Corp., 3375 S. Hoover Unit J, Los Angeles, CA 90007, tomholman@aol.com and Univ. of Southern California Integrated Media Systems Ctr.)

Sound on the desk top has some similarities and some differences from reproduction over other standardized sound systems such as those employed in the production of film and television programs. This work examines areas where the application of known acoustic and psychoacoustic principles may be used to produce a system that scales the experience of large-scale sound systems optimally to the desk-top environment. Among the methods employed are multichannel reproduction, direct-field dominant reproduction over a wide frequency range, optimal equalization for the conditions encountered by such systems, and wide frequency and dynamic ranges, among others to be discussed. Research topics underway include methods for head tracking and optimal control over the sound field for one listener, with the eventual goal of eliminating surround loudspeakers and replacing them with virtual loudspeakers. The current finding is that contemporary systems are excessively sweet-spot dependent and change the timbre too much to be acceptable professionally in this application, so the direction of research is to reduce these limitations to make a practical system with just three loudspeakers (left, right, and subwoofer).

11:20

2aSP7. Predicting the intelligibility of speech transmitted over nonideal channels. Christine M. Rankovic (Dept. of Speech-Lang. Pathol. and Audiol., Northeastern Univ., 133 Forsyth Hall, Boston, MA 02115, crankovi@lynx.dac.neu.edu)

Speech is the most efficient auditory code for conveying abstract information. Hence, high-fidelity speech transmission is an important goal for some multimedia applications. Problems may arise when the signal processing used to create the artificial auditory display inadvertently alters the spectral/temporal properties of the speech or when competing signals mask out portions of the speech. In cases where the distortion amounts to linear filtering and/or noise masking, the degrading effect on intelligibility can be predicted using an elaborate calculation scheme known as articulation theory [H. Fletcher and R. H. Galt, J. Acoust. Soc. Am. **22**, 89–151 (1950)]. The theory assumes that the audible spectrum can be divided into a number of bands that contribute independently to speech reception; and the band contributions are additive. Acoustical measurements are translated into their perceptual correlates via transformations derived from psychophysical experiments. These terms are combined to give the articulation index a number which is directly related to the expected percent-correct score for an articulation test. The theory will be described and examples presented. [Work supported by NIH R29 DC02127.]

11:50

2aSP8. Subjective evaluation of space discretized ray tracing. Soeren K. Olesen (Acoust. Lab., Aalborg Univ., Fredrik Bajers Vej 7 B4, DK-9220 Aalborg, Denmark, acoustics@kom.auc.dk)

The growing interest in multimedia raises questions like how to present 3-D sound to the user of a multimedia system. Such presentations should not necessarily be exact replicas of the real-life experiences; however, they should still provide a quality that convinces the user of being present in the acoustical environment intended. This paper briefly introduces an alternative method of ray tracing that indeed is simplified compared to that of the conventional ray-tracing/mirror-image approach. The method may be considered as a space discretized version of ray tracing and it is considerably faster than the latter. In order to reveal subjective differences between the two methods, a psychoacoustical listening experiment was carried out using binaural synthesis. Twenty subjects participated and in two sessions they were asked to rate coloration and reverberation, respectively, by clicking a mouse—each time listening to a music or speech sound frame. In a third session a sketch of a room and the position of the subject were shown, and the subjects were asked to indicate from what position the sound appeared. The sounds were produced using the new method and traditional ray tracing. The results from the listening experiment are presented in the paper.

12:05

2aSP9. On the relationship between the subjective preference of sound fields and the α-brain wave in both hemispheres. Yoichi Ando, Chingyao Chen, and Kazuki Nishio (Grad. School of Sci. and Technol., Kobe Univ., Rokkodai, Nada, Kobe, 657 Japan)

For the purpose of finding possible brain activity corresponding to the subjective preference of sound fields, an attempt was made to analyze the effective duration (τ_e) of the running autocorrelation function (ACF) in the α-wave range (8–13 Hz) in both hemispheres. As the sound source, a music motif (Arnold: Sinfornietta of Opus 48, a 5 s piece of the 3rd movement) was selected. When the temporal factor of the sound fields (the delay time of reflection Δt_1) is changed, the values of τ_e in the α wave in the left hemisphere correspond well with the subjective preference, but not in the right hemisphere. A long value of τ_e in the α wave indicates the preferred condition. When the reverberation time T_{sub} is changed, the values of τ_e in the α wave were much longer at the preferred reverberation time $[T_{\mathrm{sub}}]_p = 1.2$ s than those at $T_{\mathrm{sub}} = 0.2$ or 6.4 s in the left hemisphere. On the other hand, when the spatial factor of the sound fields (IACC) is changed, the right-hemisphere dominance is achieved by the analyses of the value of τ_e in the α wave ($p < 0.001$).

TUESDAY MORNING, 2 DECEMBER 1997 CALIFORNIA ROOM, 8:00 TO 11:45 A.M.

Session 2aUW

Underwater Acoustics and Engineering Acoustics: Advances in Underwater Acoustic Range Technology; Tracking

Peter J. Stein, Chair
Scientific Solutions, Inc., 18 Clinton Drive, Hollis, New Hampshire 03049-6576

Chair's Introduction—8:00

Invited Papers

8:05

2aUW1. Acoustic challenges in underwater range systems. Robert E. Janiesch (Naval Undersea Warfare Ctr., Newport, RI 02841)

Over the last decade, several changes both domestically and world-wide have caused a shift in the way underwater range systems need to be designed and developed. These changes include a decline in resources (funding) available for the systems, and a shift in the world military threat from a super power in the open ocean to a variety of threats in littoral environments. As a result of these changes, new challenges—many of them directly related to acoustic issues—are presented to the designers of underwater range systems that are utilized for both tactical training and T&E. This briefing will outline the requirements for the new range systems and present the challenges that need to be met along with the associated acoustic issues. This briefing will serve as a frame work for other presentations that will follow in this session.

8:25

2aUW2. Overview of tracking techniques and developments for naval undersea tracking ranges. Dean J. Smith (NUWC, Code 3823, Bldg. 160, 1176 Howell St., Newport, RI 02841)

The recent shift in the naval warfare climate from deep-water open-ocean scenarios to littoral waters has resulted in a need for the development of shallow-water training ranges that can augment the deep-water ranges that presently exist. The inherent characteristics of littoral environments (time variability, nonhomogeneity, reverberation, sound velocity variation, etc.) provide significant challenges for developing robust shallow-water tracking systems that can meet range operation and accuracy requirements. These requirements can only be met through advances in tracking techniques and technology. This presentation will provide an overview of traditional deep-water tracking range techniques and then focus on the challenges that shallow-water tracking environments bring. Finally, new tracking techniques that are being developed to overcome these challenges will be discussed.

8:45

2aUW3. Passive target tracking for future littoral test and evaluation ranges. Colin J. Lazauski and Jeffery P. Vuono (Naval Undersea Warfare Ctr., Code 3825 B104, Newport, RI 02841)

Currently, the U.S. Navy uses active acoustic pingers for tracking vessels on its test and evaluation ranges. Although tracking accuracy for these ranges is excellent, this performance comes with some penalties. For instance, test realism can be lost due to loud pinger emissions, small or prototype vehicles might require modification for pinger installation, and ship time is often lost during fitting and removal of pingers. Drawing upon the U.S. Navy's experience in passive tracking, advances in matched mode localization, and increases in computational capacity, passive acoustic tracking on T&E ranges can now be implemented. The use of pingers is not required and realism is restored to the testing scenario. Test target hardware is no longer required and the cost of the traditional shallow-water hydrophone grid can be reduced. One design for a passive tracking range utilizes acoustic arrays with vertical and horizontal apertures, *in situ* environmental monitoring, fast and accurate propagation models, and real-time tracking algorithms. Results from several recent field tests demonstrate the potentials and challenges of passive acoustic signature tracking on T&E ranges. The results are compared with ground truth data as well as range simulations and error analysis.

9:05

2aUW4. Synthetic aperture processing for underway noise ranging. Ian A. Whitehouse (108 Marine Tce, Fremantle 6160, Australia)

In early 1992, Nautronix, participated in bidding for the Underway Radiated Noise Range (URNR) tender issued by the Royal Australian Navy (RAN). The tender document stipulated a radiated noise range capable of meeting the most stringent performance requirements. After assessing all of the bids the Commonwealth declined to award a contract on the basis that none of the submissions could demonstrate an adequate figure of merit. Nautronix subsequently proposed a ranging system employing synthetic aperture principles to achieve additional processing gain and reduce the complexity of the in-water system. With funding and assistance from the RAN the program has successfully progressed through a series of discrete phases to the development of the Shallow Acoustic Underway Range (SAUR), soon to be deployed at Thistle Island in South Australia. The range will fulfill all of the performance requirements of the original URNR, with the additional benefit of localization of sources on the vessel being ranged. SAUR will be used for ongoing RAN ranging activities. Both physical and synthetic arrays are employed. The presentation will discuss some of the requirements and performance capabilities of a synthetic array and present some of the results achieved during the development process and subsequent trials.

9:25

2aUW5. Topics of acoustic research in mine countermeasures for littoral waters. Sturle Sovik and Ashild Bergh Nilssen (Forsvarets forskningsinstitutt, FFI, Norwegian Defence Res. Establishment, P.O. Box 115, 3191 Horten, Norway)

The presentation will address the status of the FFI's (NDRE) program in mine counter measure research (MCM) and the involvement of acoustic methods, equipment, and results, therein. Assessing and countering the threat of sea mines targeting surface and sub-surface vessels are strongly dependent on the signature characteristics of the vessels and the environment of the mine–ship engagement situation. Similar considerations also apply in assessing the safety of naval vessels entering or operating in a potentially mined area. Since radiated acoustic energy and the associated seismic waves are commonly used for sea-mine actuation functions, assessment of signature characterization (acoustic), the propagation characteristics of the environment, and the noise of littoral waters constitute major research and investigation tasks at FFI. The results are contributing to advances in improving the operation capabilities of the MCM force of the Royal Norwegian Navy (RNoN). The presentation will briefly outline the objectives of our research and the programs where the acoustic methods and equipment are involved. A summary of the results and current status will also be presented. Here, emphasis on the experiments of short-range measurements in support of 3-D modeling validation and modeling results will be given.

9:45–10:00 Break

Contributed Papers

10:00

2aUW6. Further advances in the development of a novel acoustic range-bearing tracking node. Peter J. Stein, Steven E. Euerle (Sci. Solutions, Inc., 18 Clinton Dr., Hollis, NH 03049), Richard K. Menoche, and Robert E. Janiesch (Naval Undersea Warfare Ctr., Newport, RI 02841)

During the next several years the U.S. Navy will be installing large area $O(1500\ \mathrm{km}^2)$ shallow-water training ranges off the U.S. mainland. These ranges require underwater acoustic tracking of targets to an accuracy of 100 m. Tracking will be performed via high-frequency pingers (10–50 kHz) located on the targets and bottom-mounted hydrophones (nodes). Development progress on a novel acoustic range-bearing node was presented by the authors at the December 1996 ASA Conference (Hawaii). The node consists of a compact concrete structure containing a minimum number of sensors (4) and implements an axisymetric acoustic horn. Further advances in the development of the device are presented here. Additional results of data processing from testing of the prototype device and results of in-air acoustic tests will be shown. A second prototype, which is currently being constructed, will be described. The design incorporates a modified horn shape, an acoustic window and support structure, and integral ''tone-pilz'' type stacked crystal transmitters located in the horn.

10:15

2aUW7. Geodetic position estimation of underwater acoustic sensors. Harold T. Vincent II (Naval Undersea Warfare Ctr. Div., Newport, RI 02841) and Sau-Lon James Hu (Univ. of Rhode Island, Narragansett, RI 02882)

An improved survey method for estimating the geodetic position of acoustic sensors placed at fixed but unknown locations on the seafloor is developed. Bottom-mounted sensors are surveyed using an extension of the classic trilateration terrestrial survey technique, i.e., making ranging measurements from reference points to the point to be surveyed. For acoustic sensors, these ranging measurements are obtained by transmitting

an acoustic signal from a near surface projector and making corresponding timing and/or position measurements. Historically, sensors have been surveyed either relatively or geodeticly. A crucial calculation required for both cases is the effective sound speed (ESP). To avoid systematic (nonrandom) errors in the ESP calculation, an unbiased measurement for the sound speed profile (SSP) is a prerequisite. However, a typical measured SSP may have a 2-m/s bias error. Consequently, the traditional methods may suffer significant errors on estimating the sensor coordinates. The newly developed method considers the timing and sound speed biases, in addition to the sensor coordinates, as parameters to be estimated. In modeling this way, the bias errors are precluded from propagating to the sensor coordinates. Results with both simulated and actual field data show that this new method has significant improvement over traditional methods.

10:30

2aUW8. Telemetry and tracking methods for noisy multipath environments. Raymond J. Kraszewski (Naval Undersea Warfare Ctr. Div., Keyport, 610 Dowell St., Keyport, WA 98345, rayk@kpt.nuwc.navy.mil)

Abstract underwater tracking of submerged vehicles in a noisy acoustic environment presents significant performance challenges caused by the presence of countermeasures, multipath interference, and propagation loss. The challenge is exacerbated in shallow water. The Naval Undersea Warfare Center (NUWC), Keyport, has assessed several techniques for improving underwater tracking system performance. The study showed that implementation of a ''telemetric'' system based upon signal modulation techniques could afford considerable benefits in performance. This has been confirmed through experiments conducted at a shallow-water tracking range site on the continental shelf area of the Washington State coast. The quality of ''raw'' tracking data rivaled that of traditional data which had undergone postprocess ''smoothing.'' To perform these experiments, a software-controlled, processor-based tracking system known as Counter Measure Tolerant Telemetry and Tracking, or CMTTT, was developed employing phase and frequency shift keyed modulation and adaptive normalization techniques to afford improved localization accuracy under conditions, which include the presence of countermeasures in shallow water. A description of the hardware configuration of CMTTT as well as the theoretical foundations for the core signal processing algorithms are presented. Test results and accuracy issues are discussed together with suggestions for further improvement of performance.

10:45

2aUW9. Acoustic modeling of narrow waters. Ashild Bergh Nilssen (Forsvarets forskningsinstitutt, FFI, Norwegian Defence Res. Establishment, P.O. Box 115, 3191 Horten, Norway)

The problem of sound propagation in narrow waters can readily be related to room acoustic problems. Different methods from room acoustics were investigated for a possible application to the underwater acoustics of fjords and straits. The frequency range of interest is 1 Hz–100 kHz, and the sound field at ranges to ~1000 m from the source is computed. For the very low-frequency range both near-field and far-field propagation must be considered. In the very-low-frequency range, 1 to ~50 Hz (possibly 500 Hz), finite-element methodology is applied. In the medium-frequency range ~50 Hz to ~10 kHz, room acoustic methods are applied. In the high-frequency range ~10 to 100 kHz, conventional underwater acoustic methods are used. The room acoustic method chosen is a hybrid model combining image source modeling and ray tracing, and with the most important quality; a diffusivity parameter of the boundaries, which makes it unnecessary to know the exact geometry of the area modeled. However, the diffusivity parameter cannot be computed, it must be assessed from experimental data. Numerous experiments have been carried out in several fjords in western Norway to validate and optimize the methods.

11:00

2aUW10. A tonpilz projector for use in an underwater horn. Mark B. Moffett, James M. Powers, and Michael D. Jevnager (Code 2131, Naval Undersea Warfare Ctr., Newport, RI 02841)

Tonpilz transducer head piston diameters are usually about one-half wavelength in diameter, so that ''rho-c'' radiation loading will be approximately achieved. However, for this underwater tracking application, the transducer is required to fit inside the throat of a 360-deg horn. The piston diameter is limited to 1 in., or about 0.22 wavelength at the resonance frequency, 13 kHz. The design is based on a lumped-element model. The PZT-8 ring stack is 2 in. long with an outer diameter of 0.5 in. and an inner diameter of 0.25 in. The head mass is 0.25-in.-thick aluminum, and the tail mass is tungsten, with a mass of 86 g. The transducer is isolated from its cylindrical housing with a syntactic foam spring. An O-ring piston seal is used, and a similar O-ring provides support for the tail mass. The results of free-field measurements will be presented and compared with the model, which will then be used to predict the performance after installation in the horn. [Work supported by Naval Air Systems Command.]

11:15

2aUW11. Underwater impulsive acoustical generation from steam bubbles. Jack Y. Dea (NCCOSC, RDTE DIV. 711, San Diego, CA 92152-6187)

One of the most common techniques to generate impulsive sound underwater is through the use of sparkers. The sound is generated through high-voltage arcing across two electrodes. Research at NCCOSC, RDT&E division has shown that modest sound levels can be produced from electrodes operated at nonarcing voltages. In this mode the main products in the electrode spacing are steam bubbles. The expansion and subsequent collapse of a single steam bubble is what generates the underwater sound. The advantages to the use of nonarcing electrodes are longer electrode life, greater reproduciblity, and the use of light-weight components. The disadvantages include lower electroacoustical conversion efficiency and lower electrical power input. The electroacoustical conversion efficiencies measured reached 0.7%. Extrapolation of efficiency data show that 1% or better efficiency may be achievable. A high-speed imaging was developed to capture the history of events in the electrode spacing. Images of various bubble events are shown and interpreted. [Work supported by ONR.]

11:30

2aUW12. Array element localization for vertical hydrophone arrays in the Arctic ocean. Stan E. Dosso (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC V8W 3Y2, Canada), Barbara J. Sotirin, and Vincent K. McDonald (NCCOSC RDTE DIV., San Diego, CA 92152-6435)

This paper describes array element localization (AEL) for two vertical line arrays (VLA's) of hydrophones, which comprise part of the Spinnaker Array, a 3-D sensor array in the Arctic Ocean. The AEL system associated with each VLA monitors the location of three high-frequency acoustic receivers, positioned at intervals along the array cable, by measuring travel times from five seafloor transponders. The transponders are comprised of a controller unit at the base of the array and four remote units at ranges of approximately 500 m. The accuracy to which the AEL receivers can be located depends directly on the knowledge of transponder positions. Two steps were used to accurately determine the transponder positions. First, travel-time measurements were made to an independent set of acoustic sensors suspended just below the ice. Second, use was made of the AEL travel-time data themselves. Both data sets were inverted using an iterative regularized inversion that applied *a priori* information to overcome the nonuniqueness/instability of the inverse problem. Monte Carlo simulation of the inversion procedure indicates that the transponders and AEL receivers are localized to an accuracy of less than 1 m.

Session 2pAB

Animal Bioacoustics: C. Scott Johnson Session: Whale and Dolphin Acoustics I

Sam H. Ridgway, Chair
NCCOSC RDTE Division, D35, San Diego, California 92152-6266

Chair's Introduction—1:00

Invited Papers

1:05

2pAB1. A window into the acoustics of whales and dolphins. C. Scott Johnson (1876 Sefton Pl., San Diego, CA 92107)

The evolutionary history of dolphins and whales is revealed through extensive adaptation of their sensory systems. Their auditory capabilities were hinted at through their acoustic repertoire after development of the hydrophone. The echolocation capabilities of dolphins, first induced by McBride in 1956, were demonstrated behaviorally in 1958, but nothing was known about their biosonar signals until technology caught up. The revelation that dolphins interrogated their environment using ultrasonic sonar pulses was complemented by demonstration of ultrasonic hearing acuity and sensitivity extending over 6 oct above that of humans. Our understanding of dolphin auditory-filter shapes and receiver-operating characteristics advanced apace with development of instrumentation for underwater acoustics. These findings formed a basis for contemporary research on cetacean psychophysics, anatomy, and acoustical ecology, which continues to reveal the intricacies and elegance of the auditory systems and capabilities of dolphins and whales.

1:25

2pAB2. The slippery slope of a Johnsonian ear: Natural variability versus natural loss. Darlene R. Ketten (Biol. Dept., Woods Hole Oceanograph. Inst., Woods Hole, MA 02543 and Harvard Med. School, Massachusetts Eye and Ear Infirmary, Boston, MA 02114), Patrick W. B. Moore (Naval Command, Control and Ocean Surveillance Ctr., San Diego, CA 92152-6254), Lois A. Dankiewicz (Sci. Applications Intl. Corp., San Diego, CA), Randall L. Brill, and William Van Bonn (NCCOSC, San Diego, CA)

The normal audiogram of the bottlenosed dolphin, first established by Johnson [C. S. Johnson, in *Marine Bio-Acoustics*, edited by W. N. Tavolga (1967), pp. 247–260], has been pivotal to nearly three decades of underwater echolocation research, most of which focused on the mechanisms and analytical limits of dolphin sonar. One important aspect of dolphin hearing that has escaped investigation is what are normal audiometric ranges, or, more important, how to determine whether differences seen in audiometric responses of a new animal represent normal variants or a pathologically impaired ear. In this study, radiologic data and audiograms from animals tested by Brill *et al.* (this session) were compared with results from postmortem histologic analyses of ears from dolphins with known hearing losses. Structural alterations in the ear x rays of the older, male dolphins were consistent with obstructive disease processes and demineralization of temporal bones from dolphins with conductive and progressive sensorineural loss. These findings imply that specific features of the older male's audiogram, such as the relatively flat, elevated, midfrequency thresholds and the rapid decrement of high-frequency responses, are diagnostic correlates for chronic infection and age-related retrograde neural loss. [Work supported by ONR N000149310940.]

1:45

2pAB3. Evidence of hearing loss in an Atlantic bottlenose dolphin (*Tursiops truncatus*). Randall L. Brill, Patrick W. B. Moore (NCCOSC RDTE DIV. D351, 49650 Acoust. Rd., Rm. 108, San Diego, CA 92152-6254), Lois A. Dankiewicz (Sci. Applic. Intl. Corp., San Diego, CA 92121-1578), and Darlene R. Ketten (WHOI, Woods Hole, MA 02543)

Auditory thresholds were behaviorally measured for two Atlantic bottlenose dolphins (*Tursiops truncatus*); a 14-year-old female, and a 33-year-old male. Stimuli were delivered directly to the lateral sides of the lower jaw via jawphones as opposed to free-field broadcasts. The female's audiogram clearly reflects the standard for this species [C. S. Johnson, in *Marine Bio-Acoustics*, edited by W. N. Tavolga, pp. 247–260 (1967)]. Previous thresholds for the male measured at age 26 indicated a hearing loss in the left ear of approximately 2 to 3 dB [*re:* 1 μPa] between 4 to 10 kHz, which were considered unremarkable. At age 33, the same male demonstrates distinctive losses. The right ear shows a 16–33-dB loss over 10–40 kHz, the range of best sensitivity. Above 55 kHz, the right ear is 2–3 dB more sensitive than the left. Both ears then decline to an upper frequency cutoff of approximately 70 kHz below the standard 120 kHz. Hearing losses due to age have been reported for this species [S. H. Ridgway and D. A. Carder, J. Acoust. Soc. Am. **101**, 590–594 (1997)]. The data reported in this paper suggest both uni- and bilateral hearing losses in the male which may be the result of age, impairment of the auditory system, or both.

2:05

2pAB4. Temporal resolution of the dolphin's hearing: Evoked-potential study. Alexander Ya. Supin and Vladimir V. Popov (Inst. of Ecology and Evolution, Russian Acad. of Sci., 33 Leninsky Prosp., 117071 Moscow, Russia)

Temporal resolution of the dolphin's auditory system was studied using the auditory brain-stem-evoked response (ABR) recording in paradigms as follows: (i) paired click and pip stimulation; (ii) rhythmic click and pip stimulation; (iii) sinusoidally amplitude-modulated tones; (iv) noise burst of various duration (the temporal summation test); and (v) gap in noise detection. An integration temporal transfer function was derived from the data of all the listed experiments using a model of a linear integrator with subsequent nonlinear transform to ABR amplitude. All the experiments listed above gave strictly corresponding results. The integration transfer function consisted of an initial high-level part about 0.3-ms long and the subsequent low-level part with a decay of around 40 dB per time decade. The short initial part of the integration transfer function makes it possible to respond to paired stimuli with interstimulus intervals down to 0.2 ms, to follow rhythmic stimuli of a rate of more than 2000 Hz, and to detect gaps in noise as short as 0.1 ms. The subsequent low-level part is also important; it dictates response properties at interstimulus intervals and gap durations as loud as dozens of ms.

2:25–2:45 Break

2:45

2pAB5. Temporary shift in delphinoid masked hearing thresholds. Sam Ridgway, Don Carder, Carolyn Schlundt, Tricia Kamolnick (NCCOSC RDTE Div. D35, San Diego, CA 92152-6266), and Wesley Elsberry (Texas A&M Univ., Galveston, TX)

Studies to determine the presence or absence of temporary shifts in masked hearing thresholds (TTS) in response to 1-s tones of 141–201 dB *re*: 1 μPa (*S*1 tones) were done at 3, 10, 20, and 75 kHz with four bottlenose dolphins and two white whales. Dolphins exhibited short-term changes in behavior above the following levels of *S*1 tones: 186 dB at 3 kHz, 180 dB at 10 kHz, 181 dB at 20 kHz, and 178 dB at 75 kHz. TTS levels were 194–201 dB at 3 kHz (two animals), 192–196 dB at 10 kHz (two animals), 193–196 dB at 20 kHz (two animals), and 192-194 dB at 75 kHz (two animals). One white whale exhibited no TTS at the highest level at 3 kHz while the other shifted at 198 dB. At 10 kHz the whales were in the upper range of the dolphins. The studies were completed without any harm to animal health or change in day to day behavior. [Work supported by the Office of Naval Research.]

Contributed Papers

3:05

2pAB6. Auditory frequency selectivity and masked hearing capabilities in an Atlantic bottlenose dolphin. David W. Lemonds, Whitlow W. L. Au, Paul E. Nachtigall (Hawaii Inst. of Marine Biol., MMRP, P.O. Box 1106, Kailua, HI 96734, lemonds@hawaii.edu), Stephanie Vlachos (Sea Life Park Hawaii, Waimanalo, HI 96795), and Herbert L. Roitblat (Univ. of Hawaii, Honolulu, HI 96822)

Frequency selectivity capabilities of an Atlantic bottlenose dolphin were examined by calculating critical ratios from masked hearing data. Absolute sensitivity to pure tones from 40–140 kHz was measured as a base line, and masked sensitivity was determined for the same signals masked by three levels of white noise (52, 42, and 32 dB *re*:1 μ Pa^2/Hz). Absolute and masked sensitivity were essentially constant between 40 and 120 kHz at each of the masking conditions. Sensitivity decreased approximately 100 dB per octave between 120 and 140 kHz. Critical ratios averaged across noise levels were constant between 40 and 120 kHz, averaging 26 dB. The critical ratio at 140 kHz was 42 dB. This flat trend in critical ratios between 40–120 kHz does not agree with the constant-*Q* filter-bank models used to account for earlier critical ratio and bandwidth measurements for the species [W. W. L. Au and P. W. B. Moore, J. Acoust. Soc. Am. **88**, 1635–1638 (1990)], but comparison to other cetacean masked hearing work [Johnson *et al.*, J. Acoust. Soc. Am. **85**, 2651–2654 (1989)] indicates that the trend has been recorded before, but not explicitly reported.

3:20

2pAB7. Frequency tuning in dolphins: Evoked potential study. Vladimir V. Popov and Alexander Ya. Supin (Inst. of Ecology and Evolution, Russian Acad. of Sci., 33 Leninsky Prosp., 117071 Moscow, Russia, popov@inemor.msk.su)

Simultaneous tone masking and notch masking in conjunction with auditory-evoked potential recording was used to measure frequency tuning in dolphins. Tuning measurements within a range of 64–128 kHz using auditory brain-stem response (ABR) in conjunction with tone masking resulted in *Q*(10 dB) values of 16–18. However, this paradigm was unusable to measure frequency tuning at lower frequencies because ABRs are evoked by short transient acoustic stimuli of wide frequency band. To avoid this limitation, narrow-band sinusoidally amplitude-modulated tones were used as probes, which produce evoked potentials as an envelope-following response. With this method, tuning curves were obtained within the frequency range from 11.2–110 kHz. The *Q*(10 dB) values of the obtained curves varied from 12–14 at 11.2 kHz to 18–20 at 64–90 kHz. The data obtained with notch-noise maskers confirmed the results of tone masking: being approximated by roex filter form, they showed *Q*(10 dB) values of 18–20 and ERB of 35–39. Thus no tuning overestimation due to off-frequency listening occurred in the tone-masking experiments.

3:35

2pAB8. Zones of masking around icebreakers affecting beluga whales. Christine Erbe (Univ. of British Columbia, EOS-Geophys., 2219 Main Mall, Vancouver, BC V6T 1Z4, Canada, erbe@geop.ubc.ca)

Masked hearing thresholds of a beluga vocalization in four types of noise were measured during behavioral experiments with one beluga whale. In particular, propeller cavitation noise and bubbler system noise created by an icebreaker were studied and compared to naturally occurring icecracking noise and artificially created Gaussian white noise. Critical signal-to-noise ratios, at which each noise just masked the beluga vocalization, were used for ocean sound propagation analysis. A three-point scenario was established involving the relative distances between a noise-creating icebreaker, a vocalizing beluga whale, and a listening conspecific. Maskograms are color plots visualizing the two overlapping sound fields and indicating zones of masking around the noise source. Results are that propeller cavitation noise (source level 200 dB *re*:1 μPa at 1 m) has a maximum radius of masking of 23 km. Bubbler system noise (source level 190 dB *re*:1 μPa at 1 m) masks over 15 km. The masking zone around natural icecracking noise (source level 150 dB *re*:1 μPa at 1 m) has a radius of only 10 m. This is the same for broadband Gaussian white noise with a source level of 160 dB *re*:1 μPa at 1 m.

3:50

2pAB9. Aerial and underwater hearing sensitivity of a northern elephant seal (*Mirounga angustirostris*). David Kastak and Ronald J. Schusterman (Long Marine Lab., Univ. of California, 100 Shaffer Rd., Santa Cruz, CA 95060, kastak@cats.ucsc.edu)

Aerial (100 Hz to 28 kHz) and underwater (75 Hz to 59.8 kHz) sound detection thresholds were determined for a female northern elephant seal (*Mirounga angustirostris*) using behavioral psychophysical techniques. The resulting underwater audiogram shows a typical phocid pattern, with an upper frequency limit of about 59 kHz and relatively flat sensitivity from about 200 Hz to 30 kHz. In air, the elephant seal is 10–30 dB less sensitive than other phocid species tested. Underwater sound pressure thresholds averaged about 15 dB lower than aerial pressure thresholds. The degree of adaptation to underwater hearing in pinnipeds seems to follow phylogenetic trends; however, a high degree of aquatic specialization occurs in the northern elephant seal. Because the elephant seal is apparently nonvocal underwater, the most likely explanation for this high degree of aquatic hearing specialization lies in its ability to dive regularly to depths of 500–700 m. Although the elephant seal has poor aerial hearing relative to other pinnipeds, it has compensated for this deficiency by emitting very loud aerial vocalizations associated with seismic and visual signals to facilitate effective aerial communication. [Work supported by ONR.]

TUESDAY AFTERNOON, 2 DECEMBER 1997 EL CAMINO/ADOBE ROOMS, 1:00 TO 5:10 P.M.

Session 2pAO

Acoustical Oceanography: Ambient Noise Inversions

John R. Potter, Chair
Acoustic Research Laboratory, Electrical Engineering Department, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260

Contributed Papers

1:00

2pAO1. Acoustic hot-spots and breaking wave noise in the surf zone. Grant. B. Deane (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238)

While the sound of breaking deep-water waves has received considerable attention since being identified as the dominant source of wind-driven noise in the open ocean five decades ago by Knudsen, there have been relatively few acoustic measurements of breaking surf. Here measurements of the acoustic signature of breaking surf in the very near shore region are presented. A two-hydrophone acoustic system was deployed approximately 100 m from the shoreline to measure the broadband coherence of single breaking waves. The horizontal directivity of the breaking wave noise can be inverted from measurements of the noise coherence. These inversions reveal the presence of acoustic ''hot-spots'' which follow the break-point of a wave as it moves along the beach. An interpretation of the acoustic hot-spots in terms of a wave's moving break-point and bubble plume formation will be given. [Work supported by ONR as part of the Adaptive Beach Monitoring experiment.]

1:15

2pAO2. Exploring the surf zone using ambient sound. Jeffrey A. Nystuen (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St, Seattle, WA 98105, nystuen@apl.washington.edu) and Ali R. Kolaini (Univ. of Mississippi, University, MS 38677)

Can the underwater ambient sound field be used to explore the oceanography of the surf zone? A preliminary examination of the ambient sound field from 1-100 kHz along the Scripps Institution of Oceanography Pier reveals two principal sound sources: breaking waves and snapping shrimp. While local wave breaking can be loud, higher-frequency sound levels within the surf zone and in rip currents are generally very quiet, confirming high-acoustic attenuation for frequencies as low as 2 kHz. Outside the surf zone, the sound field is dominated by snapping shrimp. These creatures live on the pilings (rather than free swimming), and thus present themselves as potential sound sources at distributed locations along the pier. Individual ''snaps'' are distinctive, broadband (3–100 kHz), loud, and can be tracked for over 30 m. Can these ''signals'' be used to probe the near-surf zone environment? [Work supported by ONR.]

1:30

2pAO3. Characteristics of low-frequency ambient sound in the surf zone. Ali R. Kolaini (Natl. Ctr. for Phys. Acoust., Univ. of Mississippi, University, MS 38677) and Jeffrey A. Nystuen (Univ. of Washington, Seattle, WA 98105)

An experiment was conducted along the Scripps Institution of Oceanography Pier to study the characteristics of the ambient sound in the surf zone. The role of the single large bubble oscillations versus the collective oscillations of bubble clouds as possible mechanisms of generation of the low-frequency sound, especially the relatively broadband peak that appears around 20 Hz in the sound spectrum, are discussed. The frequency band at which ambient sound from breakers in the surf leaks out is identified. The acoustic attenuation, which makes the surf zone extremely quiet above 2 kHz, and a possible method to use ambient sound to infer bubble size distribution in shallow water near the surf zone are discussed. [Work supported by ONR.]

1:45

2pAO4. An inversion method for determination of ambient noise due to surf. J. Paquin Fabre, James H. Wilson, Marshall D. Earle (Neptune Sci., Inc., 150 Cleveland Ave., Slidell, LA 70458), and Jack G. McDermid (Naval Res. Lab., Stennis Space Center, MS 39529)

Ambient noise due to surf has recently been empirically characterized as a source level density (SLD) per meter of beach as a function of wave height and frequency [Wilson *et al.*, IEEE J. Ocean Eng. **22**, 425–433 (1997) and J. P. Fabre and J. H. Wilson, IEEE J. Ocean Eng. **22**, 434–444 (1997)]. Surf noise depends upon both the source level characteristics of different surf types and the unique propagation conditions from the surf zone seaward. This dependence must be examined in order to develop models capable of providing accurate ambient noise levels in shallow water. An inverse method for determining ambient noise due to surf as a function of beach slope, sediment type, wave type, as well as frequency and wave height is presented. Discussions of incorporating the Navy Surf model into this surf noise prediction method are also provided. A long-term ambient noise data set will be used to validate the method for various environmental scenarios at one location. Using data from other locations, results can be extrapolated to give preliminary estimates of bathymetric

and geologic dependencies. The surf SLD as a function of surf type, etc., can be used for any shoreline. This makes it possible to use this model as an inverse noise model to estimate surf conditions on a beach remotely from noise measurements taken offshore.

2:00

2pAO5. Ambient noise maps of the near shore region. Marc P. Olivieri, Stewart A. L. Glegg, and Robert K. Coulson (Ctr. for Acoust. and Vib., Dept. of Ocean Eng., Florida Atlantic Univ., Boca Raton, FL 33431)

A broadband sparse volume array has been developed to study ambient noise in the near shore region [J. Acoust. Soc. Am. **100**, 2732(A) (1996)]. The array has been deployed along the coast of Boca Raton, FL, and it was found that, in the absence of boat traffic, biological noise sources were dominant and were concentrated in clusters around man-made structures or distributed over natural reefs. The sonar was deployed at different sites to triangulate on the location of these sources. Ambient noise maps of the coastal region were generated by combining the contribution from each source with a propagation model. These maps show the expected average sound pressure level from these biological sources as a function of depth and range in the along-shore and off-shore direction. Single hydrophone measurements have been made at different locations in the region to verify the maps. [Work supported by ONR.]

2:15

2pAO6. Ambient noise imaging simulation using Gaussian beams. Mandar A. Chitre and John R. Potter (Acoust. Res. Lab., Elec. Eng. Dept., NUS, 10 Kent Ridge Crescent, Singapore 119260, http://arl.nus.edu.sg)

Ambient noise imaging at high frequencies (10–100 kHz) and over short ranges (10–100 m) has been successfully simulated in the past using conventional ray tracing, combined with a Helmholtz–Kirchhoff integral approach to evaluate scattering from relatively large targets. This worked well for the design of ADONIS, the first Acoustic Daylight camera, but lacked realistic diffraction and surface roughness effects which were thought to be important in some scenarios. Furthermore, data acquired using ADONIS and other coastal ambient noise measurements have indicated that more attention should be paid to snapping shrimp as sources in warm shallow water. These sources have a rich spatial and temporal statistical behavior which have been exploited in a new generation of imaging algorithms, and now designing a second-generation ambient noise imaging (ANI) system is possible. We have an improved simulation algorithm which has therefore been developed using Gaussian beams, to investigate the feasibility and performance of a second generation ANI camera. The new simulation includes diffractive corrections and rough surface effects. The results confirm that ANI cameras can be expected to work well in warm coastal environments out to distances of at least several hundred meters. The algorithm design and some simulations will be presented.

2:30

2pAO7. Could marine mammals use ambient noise imaging techniques? John R. Potter (Acoust. Res. Lab., Elec. Eng. Dept., NUS, 10 Kent Ridge Crescent, Singapore 119260, http://arl.nus.edu.sg), Elizabeth Taylor and Mandar Chitre (NUS, Singapore 119260)

Ambient noise imaging (ANI) between 8–80 kHz has been demonstrated to produce good images at 40-m range using a 3-m aperture. With the two classic forms of sonar, active and passive, marine mammal biosonar has always outperformed man-made sonar of comparable dimensions. It is then natural to ask whether marine mammals may also use this newly discovered ANI technique as part of their sensory arsenal. Twenty odontecetes, for example, may find advantage in not broadcasting their imminent presence in pursuit of an intelligent prey. An odontecete has been observed successfully pursuing live prey without vision and with no echolocation clinks detected on any of several monitoring hydrophones during pursuit. This surprising observation may be explained if biological ANI is a viable option. To test this idea, an ANI simulation model has been adapted to include realistic target responses of fish with swimbladders, and a model of the receiving performance of a tursiops dolphin has been used to evaluate the possible performance envelope of a biological ANI system. It is found that it is plausible that some marine mammals use ANI to augment information from other senses at short ranges. The implications for understanding these animals and their behavior are substantial.

2:45

2pAO8. Where the ''Acoustic Daylight'' analogy breaks down. Nicholas C. Makris (MIT, Cambridge, MA 02139)

To make a scientific analogy between imaging with ocean-acoustic ambient noise and imaging with daylight, one must preserve the ratios of relevant physical scales. For example, the roughly meter-scale objects to be imaged by the ''Acoustic Daylight Ocean Monitoring System'' of Buckingham *et al.* [Sci. Am. (February 1996)] are millions of wavelengths across optically, but only tens to hundreds of wavelengths across in underwater sound even at up to 100 kHz. People commonly see meter-scale objects in normal diffuse daylight. When imaged with underwater sound, however, meter-scale objects behave as airborne dust particles behave in real daylight. This is because the dust particles are tens to hundreds of wavelengths across optically. Everyday, experience teaches us that dust particles are invisible in normal diffuse daylight and only become visible in special cases where they are cross illuminated by highly directional light beams. This is due to the heightened effect of diffraction and forward scattering for objects with diameters so close to the wavelength. The ''acoustic daylight'' analogy also breaks down because the signal-to-noise or mean-to-standard-deviation ratio, which sets image stability, is millions of times larger for typical optical systems in natural light than that possible in any ocean-acoustic ambient-noise imaging system.

3:00–3:15 Break

3:15

2pAO9. Feasibilities of long-range acoustic vision in the northwest Pacific with large aperture arrays. Igor N. Didenkulov, Vitaly A. Zverev (Inst. of Appl. Phys., 46 Ulyanov St., Nizhny Novgorod, 603600, Russia), and Yakov S. Karlik (Central Res. Inst. ''Morfizpribor,'' St. Petersburg, 197376, Russia)

The feasibilities of the usage of large aperture stationary arrays for the acoustic observation of the northwest part of the Pacific ocean existing near the Kamchatka peninsula are described. Such an array has a plane aperture filled with 2400 hydrophones and allows the monitoring of ocean noise distribution in a horizontal view sector of 120 deg within a frequency range up to 1400 Hz. Reconstruction of ocean noise sources distribution over the ranges up to about 1000 km from a shore can be made by a method of emission acoustic spectrotomography. Spatial resolution can be improved by the incorporation of several arrays for observation. Acoustic large-scale vision may be applied for the monitoring of wind distribution, in particular storms.

3:30

2pAO10. The influence of rainfall at the ocean surface. Jeffrey A. Nystuen (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, nystuen@apl.washington.edu)

The underwater ambient sound field can be inverted to detect, classify, and quantify rainfall at the ocean surface. This application of the ambient sound field is explained and used to produce rainfall statistics and event-specific rainfall records over a three-month period from a deep-ocean mooring. The influence of rain at the ocean surface is explored using measurements of near-surface salinity, the surface wave and wind fields, and subsurface acoustic backscatter. Enhanced injection of bubbles into the mixed layer by rain in the presence of high wind is indicated by changes to the shape of the underwater sound spectrum. [Work supported by ONR and NSF.]

3:45

2pAO11. Vertical coherence of ambient noise in shallow water: Application to broadband inversions. Michael J. Buckingham and Grant B. Deane (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0213)

Broadband (200-Hz to 20-kHz) measurements of the vertical coherence of ambient noise have been made in two well-surveyed, shallow-water channels with fluidlike, sedimentary seabeds. The noise at one of the sites, the StrataForm natural laboratory off Eureka, northern California, was occasionally found to be almost exclusively wind generated. Under such conditions, the theoretical coherence, computed using the known properties of the sediment at Eureka, closely matches the data. Subtle effects associated with the depth of the sources (bubbles from breaking waves) are evident above 6 kHz in both the theory and the data. At the second site, in Jellicoe Channel, New Zealand, wind noise was also a major factor but shipping sometimes contributed significantly to the field. Again, theory and data matched well when the known geoacoustic properties of the bottom were used as inputs to the model. Based on these results, it appears that the vertical coherence of ambient noise in shallow water is predictable over a bandwidth as high as 20 kHz, and accordingly has potential as the basis of an inversion technique for obtaining the geoacoustic properties of the sea floor.

4:00

2pAO12. Source depth estimation using modal structure extracted from vertical array ambient noise data in shallow water. Paul Hursky, William S. Hodgkiss, and William A. Kuperman (Marine Phys. Lab., Scripps Inst. of Oceanogr., San Diego, CA 92093-0701)

Modal structure is embedded in shallow water ambient noise [Kuperman and Ingenito, J. Acoust. Soc. Am. **67**, 1988–1996 (1980)]. Previously, it was reported that ambient noise measurements (SWellEx-1) using a vertical array were used to extract mode shapes [Hursky, Hodgkiss, and Kuperman, J. Acoust. Soc. Am. **98**, 2971(A) (1995)]. Further analysis has characterized source distributions which would support such inversions for mode shapes. Previously, it has been shown that mode shapes alone provide sufficient information to estimate source depth [E. C. Shang, J. Acoust. Soc. Am. **77**, 1413–1418 (1985) and T. C. Yang, *ibid.* **82**, 1736–1745 (1987)]. Several depth estimators based on mode shapes extracted from ambient noise were tested on towed source data (SWellEx-96). Results of processing simulated and experimental data will be presented to assess feasibility of source depth estimation without prior environmental measurements or modeling.

4:15

2pAO13. Matched-field localization for Arctic ice-ridging events. Stan E. Dosso (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC V8W 3Y2, Canada), Michael V. Greening (Cooperative Res. Ctr. for Robust and Adaptive Systems, Salisbury, S.A. 5108, Australia), and Pierre Zakarauskas (Univ. of British Columbia, Vancouver, BC V5Z 3N9, Canada)

This paper describes an inversion of ambient noise fields measured in the Arctic in an attempt to determine the spatial distribution of ice-ridge building events in the polar ice pack. To this end, a matched-field algorithm is developed to localize multiple broadband sources in a range- and azimuth-dependent environment where the bathymetry is poorly known. Both the source locations and unknown bathymetric parameters are included as variables in a simulated annealing inversion that searches for the best match between the measured and modeled fields. A number of aspects of this problem are considered in a synthetic study including the importance of simultaneous inversion, the effect of bathymetry errors, and the problem of determining the number of sources present. The inversion accurately localizes multiple synthetic sources. The method is also applied to a set of ambient noise measurements recorded in the Lincoln Sea in the Canadian high Arctic. The results indicate that a small number of ice-ridging events is sufficient to accurately model the measured fields. Estimates of the number of sources and source bearings are unambiguous; however, source ranges could not be estimated unambiguously.

4:30–4:40 Break

4:40–5:10

Panel Discussion

TUESDAY AFTERNOON, 2 DECEMBER 1997 FORUM ROOM, 1:30 TO 4:00 P.M.

Session 2pEA

Engineering Acoustics: Acoustic Materials, Transducers, Arrays

Thomas R. Howarth, Chair
Naval Research Laboooratory, 4555 Overlook Avenue, S.W., Washington, DC 20375-5350

Contributed Papers

1:30

2pEA1. Fabrication of low-sound-speed particulate composites for acoustic applications. Molly Frohlich, Gregory Kaduchak, and Charles M. Loeffler (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

Low-sound-speed elastic materials are of great interest in a wide range of applications. The present research addresses the fabrication of a low-sound-speed elastic material in the form of a particulate composite. A polymeric matrix material is combined with highly compressible particulates (40-μm diameter). At frequencies well below the monopole resonance of the particles, dilatational sound speeds as low as 400 m/s have been measured in experiments. The attenuation characteristics of the material are highly dependent upon frequency and may display transmissive or absorptive properties. The measured sound speeds and attenuations compare well with theoretical predictions for composites with relatively low volume fractions of particulates. Applications of these materials will also be discussed. [Work sponsored by the ARL:UT Independent Research and Development Program.]

1:45

2pEA2. Improving loudspeaker performance for active noise control applications. Steven A. Lane and Robert L. Clark (Dept. of Mech. Eng. and Material Sci., Duke Univ., Durham, NC 27708-0302)

Actuator performance is a critical part of active noise and acoustic control. The loudspeakers that are normally used as actuators in many active noise and acoustic control applications add significantly to the dynamics of the control loop and are detrimental to the controller's performance. By compensating a loudspeaker with a technique similar to motional feedback, the loudspeaker performance is enhanced in applications such as control of acoustic enclosures. In this work, a method to easily and reliably compensate a loudspeaker in order to approximate constant volume velocity behavior over the piston-mode frequency range is presented and demonstrated. This reduces the influence of the environment upon the actuator, reduces low-frequency distortion of the speaker, and minimizes magnitude and phase shifts of the intended control signal. Numerical simulations and experimental results of the proposed methods are included. [Work supported in part by the Lord Corporation and the National Science Foundation under Grant No. CMS 95-01470.]

2:00

2pEA3. Progress in microfabricated ultrasonic transducers. Igal Ladabaum, XueCheng Jin, and Butrus T. Khuri-Yakub (E. L. Ginzton Lab., Stanford Univ., Stanford, CA 94305, igal@stanford.edu)

Microfabricated ultrasonic transducers (MUTs) are generating interest as a novel alternative to piezoelectric transducers in many applications. This paper introduces the latest results, both experimental and theoretical, pertaining to the transducers the authors invented. Significant experimental results include air-coupled aluminum through transmission at 2.3 MHz with a system of 110-dB dynamic range, as well as fluid immersion transmission experiments with a dynamic range of at least 100 dB. Air-coupled transmission scans of defects in Plexiglas are demonstrated. Improvements to the authors' fabrication process are highlighted and the paper concludes with a discussion of the energy dissipation mechanisms of the transducer. The main dissipative mechanism is, of course, radiation, but the effects of structure leakage, viscous damping, and thermal flow are introduced into the authors' model. [Work supported by the U.S. Office of Naval Research and by a fellowship from NUS.]

2:15

2pEA4. Calculations of the elastic compliance and piezoelectric constants of PVDF and P(VDF-TrFE) crystals using molecular mechanics. Harold C. Robinson (Naval Undersea Warfare Ctr., Newport, RI 02841), George J. Kavarnos (Anal. and Technol., Inc., North Stonington, CT 06359), and Robert W. Holman (Western Kentucky Univ., Bowling Green, KY 42101-3576)

The selection and/or design of projector materials for ultrasonic transducers requires knowledge of their elastic, piezoelectric, and dielectric properties. These properties can be measured directly; however, in order to minimize the fabrication and measurement of many different materials, it would be useful to compute these properties from the chemical composition and structure of the material. In molecular mechanics, a force field is used to describe short- and long-range interactions between atoms in a crystal. The most favored state occurs when the total energy, i.e., the sum of the interaction energies, is minimized. This paper presents calculated compliance and piezoelectric constants for β-phase crystals of polyvinylidene fluoride (PVDF) and its copolymer with trifluoroethylene, P(VDF/TrFE). The constants were derived by minimizing the energy of the unit cell while varying the stress. The model was validated using experimental and theoretical values for the properties of PVDF. The behavior of the compliance and piezoelectric constants of P(VDF/TrFE) are examined as a function of the molecular percentage of VDF. It was found that an optimum value for d_{33} exists between 55 and 65 mol % VDF. The implications of these results for the use of P(VDF/TrFE) copolymer in ultrasonic projectors will be discussed.

2:30

2pEA5. Audio application of the parametric array. David T. Blackstock (Dept. of Mech. Eng. and Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029, dtb@mail.utexas.edu)

''HyperSonic Sound'' is American Technology Corporation's commercial application of the parametric array to produce broadband audio sound (Business Week, 12-2-96 issue, pp. 108–109). E. G. Norris demonstrated the device at the Penn State ASA Meeting [J. Acoust. Soc. Am. **101**, 3072(A) (1997)], but gave few technical details. An analysis of the device demonstrated is provided here. The primary fields are produced by sixty 40-kHz elements mounted on a ring, 4-cm i.d., 7-cm o.d., which produce an axial sound-pressure level of 140 dB at 30 cm. The values given imply a Rayleigh distance of about 1.2 m and a (primary) beamwidth of about 3°. Estimated by the absorption length for the primaries, the interaction region is about 9 m. Those who witnessed Norris's demonstration can attest to the narrowness of the secondary sound (a synthesized musical melody) and apparent absence of sidelobes. However, the melody heard also exhibited distortion. Distortion may be predicted by recalling that when the primary wave is a modulated carrier, the secondary sound is (asymptotically) the second derivative of the modulation envelope squared. Prospects for reducing the distortion are discussed [see Yoneyama *et al.*, J. Acoust. Soc. Am. **73**, 1532–1536 (1983)].

2:45–3:00 Break

3:00

2pEA6. On the use of injection molded 1–3 piezocomposite for parametric mode sonar projectors. Kim C. Benjamin, Mark B. Moffett, Kirk E. Jenne (Naval Undersea Warfare Ctr., Newport, RI 02481), and William L. Konrad (Omni Technol., Inc., Kenner, LA 70065)

Recent experimental studies have indicated that injection molded 1–3 piezocomposite is suited for parametric mode sound generation. Using a new Navy calibration standard (F82), which relies on the 1–3 composite for the active layer, the authors were able to drive the projector at sufficient power levels and measure difference frequency sound fields comparable to conventional diced PZT-based transducer designs. Although the particular 1–3 composite used in the experiment had a low primary mode efficiency, (10 μPa) were realized at a difference frequency of 50 kHz for very modest electric fields (1.5 V/mil). The subject presentation will describe the experimental results as well as include a discussion of modeling issues, and related predicted results. Suggestions will be made for improving the injection molded composite efficiency. [Work supported by NRaD Code D744.]

3:15

2pEA7. Long-term operation tests of giant magnetostrictive driving units for low-frequency sources. Toshiaki Nakamura, Iwao Nakano (Japan Marine Sci. and Technol. Ctr., 2-15 Natsushima-cho, Yokosuka, 237 Japan, nakamurat@jamstec.go.jp), and Tomohiro Tsuboi (Oki Electric Industry Co., Ltd., Numazu, 410 Japan)

A sound source with the frequency below 70 Hz and transmitting level over 195 dB is needed to investigate global ocean climate change by acoustic monitoring like the ATOC project for long-range sound propagation of 5–10 Mm of gyre scale. Long-term operation of the source over 10 years is also required for this purpose. A low-frequency sound source of 70 Hz is planned using the same driving units of giant magnetostrictive type as the units used in the 200-Hz source for our tomography system [Nakamura *et al.*, J. Acoust. Soc. Am. **96**, 3256(A) (1994)]. Acceleration tests of long-term operation of the driving units were conducted to realize the equivalent operation to 10 years of the 70-Hz source in a shorter period. As the resonance frequency of the driving unit is about 1 kHz, the acceleration rate is 1000/70. The driving unit was operated 40 000 times for 1 min, which corresponds to a 30-min operation for the 70-Hz source

every 4 h for 10 years. Variation of strain, impedance, and displacement-current sensitivity of five driving units were measured every 10 000 times operations. It was found there was no significant variation for these parameters during the experiment.

3:30

2pEA8. Analysis of mutual interaction on sonar array and vibration velocity measurement using NAH. Tomoki Yokoyama, Mitsuru Henmi, Akio Hasegawa, and Toshiaki Kikuchi (Dept. of Appl. Phys., Natl. Defense Acad., 1-10-20 Hashirimizu, Yokosuka-shi, 239 Japan, tyokoyam@cc.nda.ac.jp)

Recently, sonar has tended to use lower frequencies and large output power to enhance its search distance. However, for such sonar, it is known that unpredictable deterioration of output and the destruction of specific elements may occur because of interference between elements. This research was conducted to analyze the interference between elements in an array in which the elements were driven by a regulated voltage—the influence of the mutual radiation impedance on the vibration velocity of the elements—using an equivalent circuit. The vibration velocity of each element was calculated using mechanical parameters for a practical array element. After that, the vibration velocity regulations of elements of (1,3) and (2,2) mode array caused by interactions were verified experimentally using a NAH imaging system which makes it possible to visualize fine regulations. These results are according to simulations.

3:45

2pEA9. Dynamic mechanism of micromachined electrostatic airborne ultrasonic transducers. Li-Feng Ge (Dept. of Phys., Anhui Univ., 3 Feixi Rd., Hefei 230039, China)

Micromachining technology allows electrostatic (or capacitive) airborne ultrasonic transducers to be made with highly reproducible characteristics, thus providing the possibility to model their behaviors reliably. The behavior of a transducer depends on its geometric structure and the materials to be used. Its mechanical stiffness results from the compressibility of air in cavities, the bending stiffness of the diaphragm, and the plane tension applied to the diaphragm, if any. It is clarified that there are two main types of dynamic mechanisms according to different structures. In one type in which the air in cavities is enclosed, the diaphragm can be treated as a thin plate supported by an air spring or a plate founded on an air cushion, thus adopting the plate-on-air–spring model or short-tube model, such as those transducers with V-grooved, U-grooved, or pit-array-texture backplates [L.-F. Ge, Chin. Sci. Bull. Acad. Sin. **10** (1997)]. In a second type, air is not enclosed, so that the effect of air spring can be omitted, thereby taking the membrane-under-tension model, such as conventional condenser microphones and ultrasonic transducers with acoustic holes. Resonant frequencies of a transducer can be determined by the vanishing of the reactance of its inversive impedance [L.-F. Ge, J. Acoust. Soc. Am. **96**, 3318 (1994)]. [Work supported by the 211 Engineering Foundation of Anhui University.]

TUESDAY AFTERNOON, 2 DECEMBER 1997 COUNCIL ROOM, 1:15 TO 4:35 P.M.

Session 2pMUa

Musical Acoustics: Piano

Uwe J. Hansen, Chair
Department of Physics, Indiana State University, Terre Haute, Indiana 47809

Chair's Introduction—1:15

Invited Papers

1:20

2pMUa1. Acoustics of piano tone. Daniel W. Martin (7349 Clough Pike, Cincinnati, OH 45244)

A tutorial review of research into acoustics of piano tone emphasizes and includes tape demonstrations of the importance of tonal attack, envelope slope, optimum tuning, voicing technique, and directivity. Recent research articles about structural and dynamic piano properties and their effects upon piano tone are surveyed.

1:50

2pMUa2. Large grand versus small upright pianos: Factors of timbral difference. Alexander Galembo and Lola L. Cuddy (Dept. of Psych., Queen's Univ., Kingston, ON K7L 3N6, Canada, galembo@psyc.queensu.ca)

The difference in timbre, especially in the bass range, is one of the important reasons why pianists prefer large grand pianos to small uprights. The work discussed in this presentation addresses acoustical and design factors responsible for this difference. [Work supported by NATO and NSERC.]

2:20

2pMUa3. Impact excitation of piano string vibration. Donald E. Hall (Phys. Dept., California State Univ., 6000 J St., Sacramento, CA 95819)

Piano strings are well described by linear partial differential equations, even when stiffness is included. But compressed-felt piano hammers have a distinctly nonlinear force–compression relation, so it is challenging to model the hammer–string interaction correctly. In recent years it has been practical with personal computers to carry out numerical integration of the nonlinear equations for specific examples, though hammer/string mass ratio, hammer compliance, hammer resistive loss, string stiffness, striking point, and

striking velocity present a large parameter space to explore. Examples of such integrations, showing calculated string motions and spectra, will be presented on videotape. Related investigations have been carried out by several authors, and are of interest both for offering an understanding of energy transfer to the strings in acoustic pianos and for suggesting digital algorithms for synthesizing sounds that represent a generalization of the piano family.

2:50–3:05 Break

3:05

2pMUa4. Piano bridge and soundboard motion. N. Giordano (Dept. of Phys., Purdue Univ., West Lafayette, IN 47907, ng@physics.purdue.edu)

The bridge of a piano is the place where two important components, the strings and the soundboard, meet and interact. The motion of the bridge is important for understanding soundboard vibrations and sound production. This motion is also central to the interaction between strings, and to coupling of the different modes of a single string. Recent experimental and theoretical studies of bridge and soundboard motion are described. While the dominant motion is, as one would expect, perpendicular to the plane of the soundboard, the experiments show that motion parallel to this plane is also quite significant. Quantitative results for the in-plane motion are presented, and its role in sound production is discussed. We also consider the implications for modeling of soundboard and string motion. [Work supported by the National Science Foundation through Grant PHY-9722031.]

3:35

2pMUa5. Correlation of sound board normal modes with the radiated sound field of a grand piano. Uwe J. Hansen (Dept. of Phys., Indiana State Univ., Terre Haute, IN 47809) and Ingolf Bork (Physikalisch-Technische Bundesanstalt, D-38116 Braunschweig, Germany)

An extension of experimental modal analysis techniques enables the representation of the radiated sound field of a grand piano in reference to the phase of the excitation on the bridge. Sound field representations are illustrated for locations directly above and below the sound board and, in addition, in a plane directly in front of the instrument. The frequencies are chosen to coincide with the lower normal mode frequencies of the sound board, and the animated visual representations of the sound field are juxtaposed for comparison with modal analysis animations for the normal modes of the sound board.

4:05

2pMUa6. Reproducing a piano performance. George F. Litterst (Preparatory School, New England Conservatory of Music, 28 Daniel St., Newton Centre, MA 02159-2502) and Mike Bates (Yamaha Corp. of America, Buena Park, CA 90620)

There are many ways to store a piano performance for later reproduction, each of which lacks a complete and consistent definition. A composer typically notates a score for a pianist to play from—leaving many decisions to the whim of the performer. A pianist practices and commits hand and foot movements to memory—none of which produces a consistent result. A recording engineer captures the vibrations of sound for reproduction on a variety of playback systems—each of which plays back with individual characteristics which, from the listener's point of view, depart significantly from the experience of the live performance. The Disklavier piano offers a different approach which is designed to capture and play back a performance accurately by recreating the original performance on an actual piano. The information which drives the disklavier can be output to a computer in order to display performance data and to facilitate an analysis of the performance.

TUESDAY AFTERNOON, 2 DECEMBER 1997 COUNCIL ROOM, 4:40 TO 6:00 P.M.

Session 2pMUb

Musical Acoustics: Demonstration of a Computer-Based Piano Performance Reproduction System

Uwe J. Hansen, Chair
Department of Physics, Indiana State University, Terre Haute, Indiana 47809

Chair's Introduction—4:40

In this session a Yamaha Disklavier Grand Piano computer based performance reproduction system will be demonstrated. The instrument incorporates optical sensing devices which monitor onset of key motion and hammer velocity. This information is recorded and then used to drive electromagnetic plungers to activate the piano keys in a dynamic performance reproduction on the original instrument. Computer control permits editing as well as transposition and other technical modifications. The computer connection also enables the use of the instrument as a MIDI input or output device. The demonstration will include reproduction of a number of performance examples by noted artists. The demonstration will take approximately 75 minutes. It will be repeated on Wednesday at 8:30 A.M. The instrument will also be available, by arrangement, for small group hands-on evaluations.

Session 2pNSa

Noise and Architectural Acoustics: Progress Report and Discussion on the Continuing Activity on ASA's Role in Noise, Its Effects and Its Control

T. James DuBois, Cochair
DuBois & Associates, 9424 Crystal View Drive, Tujunga, California 91042

David Lubman, Cochair
D. Lubman & Associates, 14301 Middletown Lane, Westminster, California 92683

1:00

The Task Forces of several Technical Committees are involved with outreach activities related to the effects of noise and its control. This session will discuss their progress. Members of the Task Force of the Technical Committee on Noise will review the results of a successful special session on meeting-room acoustics presented to the Council of Engineering and Scientific Society Executives (CESSE) in Pittsburgh in July 1997. Members of a joint Task Force of the Technical Committees on Architectural Acoustics, Noise, and Speech Communication will review the results of a well-received seminar on classroom acoustics that they presented to Los Angeles Unified School District (LAUSD) Architects in July 1997. This Task Force, augmented by support from the Technical Committee on Psychological and Physiological Acoustics, is preparing a workshop on ''Eliminating Acoustical Barriers to Learning in Classrooms'' to be held in Los Angeles on December 6 and 7, 1997. The Task Force will discuss the necessity for the nationwide presentation of the workshop. In addition, progress will be reported on: (a) the development of a video of noise control demonstrations for use in classrooms, (b) prospects for hands-on automated hearing screening testing in public spaces, and (c) the possible exploration by ASA of opportunities for members of the American Institute of Architects to receive Continuing Education Units by participating in the December workshop and similar activities related to improved acoustics in meeting-rooms or classrooms.

Contributed Papers

2:00

2pNSa1. Measuring speech intelligibility in K–5 classrooms in a Seattle public school. Dean Heerwagen (Dept. of Architecture, Univ. of Washington, Box 355720, Seattle, WA 98195-5720), Karen Anderson (Rainer Audiol. Consulting Service, Puyallup, WA), Dale Lang (Univ. of Washington, Seattle, WA 98195-3600), and Robin Towne (Sound Judgments, Seattle, WA 98117)

A speech intelligibility pilot study is being conducted in primary school classrooms in Seattle. The properties of a number of unoccupied, conventional classrooms have been determined [e.g., EDT, RT60, and background noise levels without and with operating mechanical (HVAC) systems]. Intelligibility tests have been conducted in one room using four groups of children, two from grades 1–3 and the others from grades 3–5. The tests employed prerecorded AUDITEC words lists and Test of Auditory Comprehension short paragraphs and picture images. The test conditions used prerecorded tapes played through a loudspeaker placed at the front of the class. These tapes were played back at levels 15 dB(A) above background noise levels. Four playback conditions served as the major variable for these tests: with the HVAC system ''off;'' with HVAC system ''on;'' through an FM system; and a repeat of the first condition. Other test variables are student grade levels and student's location versus the loudspeaker. An additional variable arises from the ethnic diversity of the school population: 46 distinct ethnic groups are present in the student body. Data analysis is presently underway. [Work supported by University of Washington Royalty Research Fund.]

2:15

2pNSa2. Noise in academic environments. Sergio Beristain (Fac. of Elec. and Mech. Eng., Natl. Polytech. Inst., P.O. Box 75805, 07300 Mexico City, Mexico)

In the Acoustics Laboratory of the Faculty of Electrical and Mechanical Engineering of the National Polytechnic Institute (Mexico), a research project on the ''Evaluation of Noise in Academic Environments'' is under development. During the project, noise measurements have been carried out inside and outside of several educational installations of different grades. This paper summarizes some noise levels and describes critical conditions found in and around schools. Inside schools, the most critical cases were found among the lower educational levels, where kids became quite noisy, depending on the teacher's authority, and sometimes on the economical condition of students. In other cases, it was the community surrounding some given schools that was the source of the annoying noise. At the bachelor of science and higher levels, it can be said, in general, that noise from the school complexes to the community is irrelevant. Nevertheless, there were periods of time and conditions with high noise levels which affected academic activities. Due to the lack of national standards and recommendations for the evaluation of noise from or to schools, measuring procedures were taken from ''Labor Noise Conditions'' and ''Noise Sources'' national standards, and reference was made to international recommendations for classrooms, auditoriums, etc.

Session 2pNSb

Noise and Psychological and Physiological Acoustics: Effects of Fetal Noise and Vibration Exposure

Mary M. Prince, Chair
IWSB, NIOSH, DSHEFS, MS R13, Cincinnati, Ohio 45006

Chair's Introduction—2:45

Invited Papers

2:50

2pNSb1. Transmission of continuous and impulse noises to the fetus in utero. Kenneth J. Gerhardt (Dept. of Commun. Processes and Disord., Univ. of Florida, Gainesville, FL 32611), Robert M. Abrams, Xinyan Huang (Univ. of Florida, Gainesville, FL), and Linda L. Pierson (U.S. Total Army Personnel Command, Alexandria, VA 22332-0417)

Sounds which originate outside the abdomen of pregnant women reach the fetal inner ears after being filtered by the materials surrounding the fetal head and by its skull. In sheep, exogeneous sounds are low-pass filtered from 0.25 to 4.0 kHz at a rate of 6 dB/octave before reaching the head. There is little difference in sound pressures for low-frequency signals when recorded inside and outside the abdomen. The transmission route for fetal stimulation is via bone conduction with low-frequency sounds (<0.25 kHz) being attenuated by about 10 dB and higher-frequency sounds (0.5–2.0 kHz) attenuated by up to 45 dB. The amplitude and waveform of an impulse is influenced markedly by the location of the recording hydrophone within the uterus. Close to the intra-abdominal wall, the impulse is best characterized as a simple Friedlander wave (duration <1.0 ms) with a peak sound pressure level (pSPL) that is approximately 5 dB less than that recorded in air. With the hydrophone located deep inside the uterus, the pSPL is reduced by over 15 dB and the waveform resembles that of an impact with a decay time of greater than 12 ms. Fetal cochlear microphonic resembles the waveform of the impulse when it is recorded with a hydrophone located by the fetal head.

3:15

2pNSb2. Effects of intense noise exposure on the auditory brain-stem response and inner ear histology of fetal sheep. Linda L. Pierson (U.S. Total Army Personnel Command, TAPC-OPH-MS, 200 Stovall St., Alexandria, VA 22332-0410), Kenneth J. Gerhardt, Robert M. Abrams, and Xinyan Huang (Univ. of Florida, Gainesville, FL 32611)

Past evidence from investigations in which an intense noise was delivered to the fetal hearing mechanism has been contradictory. Some investigations suggest the possibility of a critical period during auditory development when exposure to intense noise may adversely affect auditory function. Other studies have reported no effect of intense noise on the auditory brainstem response of fetal sheep. In one recent experiment, auditory brain-stem response thresholds were recorded in utero from fetal sheep prior to and following an intense broadband noise exposure delivered at a developmental period when the ABR was emerging. The ABR thresholds recorded from the noise-exposed fetuses, when measured over developmental time, were not as sensitive as thresholds obtained from age-matched control fetuses. In a related experiment, sheep were exposed four times to an intense broadband noise and the cochleae were harvested 20 days later for histological analysis. Histological comparisons revealed significant inner and outer hair cell damage in the noise-exposed fetuses. This damage was not observed in control fetuses. These studies provide increased evidence that intense noise delivered to a pregnant ewe can result in alterations of structure and function in the developing auditory system of the sheep fetus.

3:40

2pNSb3. Challenges in designing epidemiologic studies of hearing among children exposed to noise in utero. Linda M. Frazier and Craig Molgaard (Dept. of Preventive Med., Univ. of Kansas School of Medicine-Wichita, 1010 N. Kansas Ave., Wichita, KS 67214)

An epidemiologic study by LaLande and colleagues has found hearing loss among children of women who were occupationally exposed to industrial noise during pregnancy. Other populations of noise-exposed women should be studied to determine if these findings can be replicated and to characterize the dose–response curve. There are many challenges to overcome in order to carry out such a study. These include: (1) obtaining access to an exposed population; (2) obtaining a large enough sample size, while avoiding selection bias; (3) accurately measuring noise exposure levels during pregnancy; (4) assessing the potential impact of confounders, such as noise exposure during toddlerhood, medical causes of hearing loss not related to noise exposure, and others; (5) obtaining adequate follow up of the cohort of children, and then obtaining a high consent rate to perform audiograms to detect small but clinically important decrements in hearing thresholds; (6) maintaining audiometry quality control if subjects are geographically dispersed; (7) obtaining funding for the study; and (8) developing preventive interventions that will minimize discrimination against women who work in noisy environments. Although these challenges exist, they are not insurmountable.

4:05

2pNSb4. The law against fetal protection policies for reproductive health hazards in the workplace: Implications for noise and vibration exposure. Ilise L. Feitshans (Compliance Systems Legal Group, 5355 Henry Hudson Pkwy., Riverdale, NY 10471, sljc49a@prodigy.com)

Is there a legal basis for employing special risk assessments or other preventive strategies to protect fetal health when pregnant women are exposed to potentially dangerous levels of occupational noise and vibration that may cause adverse effects on pregnancy outcome or cause hearing loss? A fetus in utero is highly susceptible to risks from toxic or hazardous agents in the workplace. Yet, complex litigation before the U.S. Supreme Court in IUAW v Johnson Controls held that fetal protection policies are illegal. The Court held that despite the risks to female employees of ''childbearing age and capability'' the law prohibits excluding women from well-paying jobs that have high lead exposure. While articulating women's equal opportunity to work in high-risk occupations during or before pregnancy, the Court did not require risk assessment or state guidelines regarding possible maternal or fetal exposures to workplace harms. This paper will explore legal underpinnings of fetal protection policies under Title VII discrimination law, the Occupational Safety and Health Act of 1970, and the related precedents concerning fetal protection to consider for the first time, the implications of these laws in workplaces where pregnant women are exposed to hazardous levels of noise and vibration.

4:30–4:35 Break

4:40–5:00
Panel Discussion

TUESDAY AFTERNOON, 2 DECEMBER 1997 CABINET ROOM, 1:30 TO 3:30 P.M.

Session 2pPA

Physical Acoustics: Inhomogeneous Media

Craig J. Hickey, Chair
National Center for Physical Acoustics, University of Mississippi, Coliseum Drive, University, Mississippi 38677

Contributed Papers

1:30

2pPA1. Wave propagation in a unique class of inhomogeneous porous materials. Craig J. Hickey, James M. Sabatier (Natl. Ctr. for Phys. Acoust., Univ. of Mississippi, University, MS 38677, craigh@sparc.ncpa.olemiss.edu), Tim J. T. Spanos, and Vicente de la Cruz (Univ. of Alberta, Edmonton, AB T6G 2J1, Canada)

A system of equations describing wave propagation in a unique class of inhomogeneous porous media is presented. The specific porous medium considered consists of an elastic matrix whose pores are filled with a viscous compressible fluid. The two components, solid and fluid, are microscopically homogeneous but the porous medium is inhomogeneous because the unperturbed porosity varies with position. Volume-averaging theorems are used to construct the general equations which form the basis for the analysis. It is shown that, in general, the presence of a porosity gradient introduces new coupling terms between the fluid and solid phases in the equations of motion. Since inhomogeneity usually implies anisotropy the problem is further specialized for the case of transverse anisotropy. It is further assumed that the wave propagates in the direction of the porosity gradient, which is a principal direction of the anisotropy. The equations are compared to recent work describing rigid framed grounds with exponential porosity profiles. [Work supported by ONR and USDA ARS National Sedimentation Laboratory.]

1:45

2pPA2. Calculation of the poroelastoacoustic response of a multilayered infinite planar porous material excited by acoustical and mechanical line sources. Stéphane Rigobert and Franck C. Sgard (LASH, DGCB URA CNRS 1652, ENTPE—Rue Maurice Audin, 69518 Vaulx-en-Velin Cedex, France, rigobert@mercurey.entpe.fr)

Today, the prediction of the poroelastoacoustic response of multilayered porous media acoustically or mechanically excited for noise control applications, is of major concern in fields such as aeronautics, the automotive industry, building acoustics, or environmental acoustics. In this talk, a technique to evaluate the two-dimensional poroelastoacoustic response of a multilayered planar porous material due to both acoustical and mechanical line sources is presented. The technique based on a Fourier wave transform takes into account propagation phenomena into both the porous material skeleton and pores. Using the equations derived from the homogeneization theory for fluid-saturated porous media, propagation constants for the $P1$, $P2$ compressional wave and the S shear wave are determined. Applying boundary conditions at the media interfaces yields a linear system of equations in terms of reflection and transmission coefficients. This allows for the calculation of displacement amplitudes of the allowed wave motions in each medium. Numerical results showing the poroelastoacoustic responses (sound pressure level above the material, mean-square velocity of the skeleton at the air–porous interface, surface impedances) of several porous materials are presented in the case of both excitations.

2:00

2pPA3. A modal reduction technique for the finite-element formulation of Biot's poroelasticity equations in acoustics. Franck C. Sgard (LASH, DGCB URA CNRS 1652, ENTPE—Rue Maurice Audin, 69518 Vaulx-en-Velin Cedex, France, franck.sgard@entpe.fr), Noureddine Atalla, and Raymond Panneton (Univ. de Sherbrooke, QC J1K 2R1, Canada)

Recently, finite-element models based on displacement–displacement and pressure–displacement formulations of Biot's poroelasticity equations have been extensively used to predict the acoustical and structural behavior of multilayer structures. These models, while accurate, lead to large frequency-dependent matrices for large finite-element models and spectral analyses problems necessitating important setup time, computer storage, and solution time. Lately, some authors have established criteria for low-frequency approximations of viscous and thermal effects that allow for a linearization of the poroelastic eigenvalue problem. Complex modes for the damped system or real modes for the undamped system can then be found and used to solve the forced problem. However, this method is not general since it requires assumptions on the porous material physical properties to meet these criteria. It is proposed in this talk to investigate another modal reduction technique. The technique, referred to as *selective modal analysis* uses a dual basis associated with the skeleton *in vacuo* and the equivalent fluid in the rigid skeleton limit, respectively. The theory behind the technique will be presented together with numerical examples illustrating its performance.

2:15–2:30 Break

2:30

2pPA4. Sound velocity measurements in partially water saturated sand. Doru Velea, James M. Sabatier, and F. Douglas Shields (Natl. Ctr. for Phys. Acoust., Univ. of Mississippi, University, MS 38677, dvelea@olemiss.edu)

The addition of small amounts of gas into a fully saturated sand drastically reduces the sound velocities in that material. The technique used to wet the sand in a series of experiments was to allow the water to be pulled into the granular material by capillary action. This process can entrap various amounts of air, depending on the grain size and the negative tension under which the water is allowed into the sand. The moisture content was measured using the time domain reflectometry method. Compressional and shear-wave velocities were measured as a function of water content in the sand. An average of 15% decrease in velocity was observed as the water content was increased from room dry state to 97% saturation. The change in velocities over this saturation range can be largely attributed to the change in bulk material density. The saturating technique used in these experiments did not achieve a high enough saturation range where a significant increase in compressional velocity has been observed. [Work supported by ONR.]

2:45

2pPA5. Effects of water vapor on acoustical properties of glass beads and Ottawa sand. Yunlong Wang, James Sabatier, and F. Douglas Shields (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677, pawang@olemiss.edu)

Shear and compressional wave velocities have been measured in glass beads and Ottawa sand as a function of the percent of the void filled with water. Air-dry glass beads and Ottawa sand were evacuated to a pressure of about 10 mTorr and then exposed to dry nitrogen and water vapor, respectively. The results show that the water vapor increased the compressional wave velocity by a factor of 2 and shear wave velocity by a factor of 3 in glass beads at the water saturation of 0.15%. For Ottawa sand, however, the change in velocities was only a few percent for the same water saturation. At present, it is proposed that vapor condenses to form a ''water lens'' at the points of grain contact. This condensation results from surface tension lowering the vapor pressure above a curved meniscus. As yet there is no satisfactory explanation of the difference in the effect of the condensed water in the beads and sand. [Work supported by the Office of Naval Research.]

3:00

2pPA6. Nonsmooth impedance profile identification using reflection data. Marcelo Magalhaes and Roberto Tenenbaum (Acoust. and Vib. Lab., Federal Univ. of Rio de Janeiro, CP 68503, 21945-970, Rio de Janeiro, Brazil)

Many works published about identification of acoustic properties of inhomogeneous media present methods based on the assumption that these properties vary smoothly along the propagation direction. This work presents a comparison between two time-domain sequential inverse methods for evaluation of either smooth or nonsmooth impedance profiles of transversely infinite media excited by plane waves at normal incidence. Sequential methods, also known as layer stripping methods, despite having the disadvantage of being quite sensitive to noise in the measured signal, are faster than other approaches, such as those that use global optimization techniques. The first method in the comparison [R. A. Tenenbaum and M. Zindeluk, J. Acoust. Soc. Am. **92**, 3364–3369 (1992)] assumes a smooth profile by using a transformation of variables, which neglects the refraction effects between two consecutive layers. The second one, more general, performs some more calculations since it takes these effects into account, paying the price of a higher CPU time, being suitable for nonsmooth profiles. Numerical experiments are carried out in order to establish applicability limits for both algorithms as discontinuities are included in the impedance profile, and also to evaluate their sensitivities when the identification is performed using noisy data.

3:15

2pPA7. Utilizing a modified surface roughness parameter for describing rough surfaces with an equivalent impedance. James P. Chambers and James M. Sabatier (Natl. Ctr. for Phys. Acoust., Univ. of Mississippi, University, MS 38677)

Previous experimental and field work has shown that rough surfaces may be modeled acoustically as though they were smooth surfaces with a modified surface impedance. In these works the roughness length scale used was a single parameter equal to 1/2 of the average height of the surface roughness elements. Recent experiments have suggested that this simple form for the roughness parameter may yield incorrect results for certain conditions such as grazing incidence over impermeable surfaces or sparsely scattered roughness elements. Modifications to this roughness parameter will be investigated. Specifically, the influence of the slope of the roughness elements, which is quantified using the standard deviation in height and the correlation length, will be investigated. [Work supported by the USDA ARS National Sedimentation Laboratory.]

Session 2pSA

Structural Acoustics and Vibration: Computational Vibroacoustics II

Noureddine Atalla, Chair
G.A.U.S., Mechanical Engineering, University de Sherbrooke, Sherbrooke, Quebec J1K 2R1, Canada

Invited Papers

1:00

2pSA1. Overview of the numerical models for the vibroacoustics of multilayer panels with porous materials. Raymond Panneton (GAUS, Dept. of Mech. Eng., Univ. de Sherbrooke, QC J1K 2R1, Canada)

The vibration and noise control of mechanical and architectural structures often involves the use of multilayer panels containing sound absorbing materials. For these applications, different types of absorbing materials may be used in different types of multilayer configurations, which will be suitable to reach the manufacturer's goals. Depending upon the type of multilayer configurations and materials to be used in a final product, the engineer has to use proper modeling tools to predict its vibroacoustic response. Commercially available codes offer the possibility of modeling sound absorbing materials in simple or complex configurations using different types of mathematical formulations. In this case, the engineer has to define what type of formulations is the best suited for his analysis. These formulations are often prohibitive whether in terms of computer time, memory requirement, or geometrical configurations. The engineer must be aware of the limitations related to each formulation. This paper presents an overview of the recent numerical models for porous media that are available for the analysis of multilayer panels. A brief description of the models and their limitations will be discussed. Also, comparisons between some of the models and experimental results will be presented.

1:30

2pSA2. Comparison of BEM and FEM for computation of acoustic resonant response in a laser chamber. Mary Baker and Paul Blelloch (Structural Dynam. Res. Corp., Ste. 200, El Camino Real, San Diego, CA 92130, mary.baker@sdrc.com)

For lasers used in the computer chip etching process, laser performance, as measured by uniformity in the energy output for etching, is affected by the density of the chamber gas at the laser electrode. The laser produces longitudinal shock waves that propagate transverse acoustic pulses. Acoustic modes of the gas within the laser chamber respond to these pulses and can create a resonant response resulting in a very nonuniform energy output of the laser. These phenomena have been studied using finite-element (FEM) and boundary-element (BEM) methods to model the laser chamber and to predict the forced response. The effectiveness of these two approaches is compared. Accuracy, computing efficiency, and ease of modeling have been considered. The challenge that is addressed is the representation of the internal structure within the chamber. The baffling effects of the internal structure provide the means to shape the response either exaggerating or attenuating the resonance. BEM and FEM methods are explored as design tools to predict the response and achieve the desired uniformity of pressure and density.

2:00

2pSA3. Development and validation of vibroacoustic models of irregular plates. Douglas D. Crimaldi and Rajendra Singh (Dept. of Mech. Eng., Ohio State Univ., Columbus, OH 43210-1107)

Irregularly shaped cover plates used in automotive gearboxes receive vibration energy generated by meshing gears, resonantly amplify and radiate sound at many frequencies. To reduce the vehicle noise levels, a metal–plastic–metal composite plate with a constrained viscoelastic layer may be utilized. Due to the complex nature of the structure, nonuniform material properties and bolted boundary conditions, it is difficult to develop vibroacoustic models. Hence, two simplified covers must be studied first: a stamped steel cover with identical geometry and nonuniform material properties and a modified flat cover which simplifies the geometry while providing uniform material properties including thickness. Each cover is studied with free and bolted boundary conditions. A new model of the bolted connection is described and implemented. Finite element and boundary element models for each cover are developed and validated based on the experimental modal analysis and sound directivity results.

Contributed Paper

2:30

2pSA4. Sensitivity analysis for vibroacoustic interior problems. Pi-Jen Kao (Structural Dynam. and Res. Corp., 2000 Eastman Dr., Milford, OH 45150) and M. A. Hamdi (STRACO, 60471 Compiegne, Cedex, France)

A semianalytical sensitivity analysis method is proposed to calculate the eigenvalue and response sensitivity for a coupled vibroacoustic system. This sensitivity information can be used in a design trade-off study or in a formal design optimization procedure. In this study, an acoustic finite-element method is used to compute the eigenvalues, eigenvectors, and responses of the acoustic interior problems. A symmetric form of the coupled equation is used here. By directly differentiating the coupled equation, it can be shown that the sensitivities of the coupled systems depend only on the eigenvalues, eigenvectors, and the derivatives of coefficient matrices of the coupled equation. These coupled sensitivities will be demonstrated for a coupled plate and box problem. Numerical results will be given for the sensitivity of elastoacoustic frequencies due to the change of plate thicknesses and the sensitivity of interior pressures due to the change of acoustic absorbing materials.

Session 2pSC

Speech Communication: Sources of Individual Variability in Speech Production and Perception

Kenneth W. Grant, Chair
Army Audiology and Speech Center, Walter Reed Army Medical Center, Research Center, Washington, DC 20307-5001

Chair's Introduction—1:00

Invited Papers

1:05

2pSC1. Individual differences in cyclical and speech movement. Joseph S. Perkell, Majid Zandipour, and Melanie L. Matthies (Speech Commun. Group, Res. Lab. of Electron., MIT, Rm. 36-591, 50 Vassar St., Cambridge, MA 02139-4307, perkell@speech.mit.edu)

A comprehensive account of speech production should explain interspeaker variation. Toward this end, a study was performed with eight speakers to investigate whether individual kinematic performance limits, as reflected in a speechlike cyclical task, could account for differences in speech kinematics. Kinematic data from cyclical CV movements at rates from 1–6 Hz of the lower lip, tongue blade, and tongue dorsum were compared with data from speech utterances in different conditions, including normal, fast, clear, and slow. There were differences in movement distance, peak velocity, and duration among the speaking conditions and among the speakers. Three speakers produced ''clear'' speech utterances with distances, peak velocities, and durations that were greater than normal. The data from two of the three may reflect some increased effort for clear speech. The amount of overlap of the speech data and cyclical data varied across speakers, ranging from little overlap to complete overlap. Thus, in general, the cyclical task did not define a kinematic performance space within which speech movements were confined. These results will be discussed in relation to differences between the tasks, possible interactions between production and perception, and possible differences in motor and/or perceptual performance capacities. [Work supported by NIDCD.]

1:30

2pSC2. Individual differences in speech production. Keith Johnson (Department of Linguistics, Ohio State Univ., Columbus, OH 43210-1298)

Common experience suggests that, to a certain extent, people can be recognized by their speech, even over the limited bandwidth of a telephone line. Such talker individuality is at least partly determined by anatomical features such as vocal-fold and vocal-tract geometry. Additionally, linguistic factors such as dialectal and speech community membership give rise to noticeable variation across talkers. In this paper two acoustic/phonetic studies which sought to hold these sources of variation constant will be presented. The first is a study of the relative strengths of (perceptually trading) acoustic cues for the /s/–/sh/ distinction in English. The talkers in this study were a dialectally homogeneous group of female college undergraduates. Substantial individual differences in the relative strengths (in a discriminant analysis) of acoustic cues such as F2 transition onset, frequency of the first fricative pole, and duration of the fricative noise were found. The second study explores the patterns of pronunciation exhibited by identical twins. Previous research suggests that identical twins have virtually identical long term average spectra, and that their voices are difficult to discriminate. The present study is designed to look in closer detail at twins' patterns of pronunciation in diphthongs, voice onset time, and cue trading.

1:55

2pSC3. Individual differences and the link between speech perception and speech production. Rochelle Newman (Dept. of Psych., 11 Seashore Hall, Univ. of Iowa, Iowa City, IA 52242, rochelle-newman@uiowa.edu)

Prior research has suggested that speakers vary in the precise muscle movements used to make phonemic distinctions [K. Johnson, P. Ladefoged, and M. Lindau, J. Acoust. Soc. Am. **94**, 701–714 (1993)]. There has also been evidence that individuals differ slightly in their perception of speech [V. Hazan and S. Rosen, Percept. Psychophys. **49**, 187–200 (1991)]. One possibility is that these sources of variability are due to the same underlying cause. In support of such a notion, data will be presented briefly demonstrating the range of acoustic-phonetic variability across individuals in both production and perception independently. Following this, evidence will be provided suggesting that these sources of variability are related. More specifically, individuals' perceptual prototypes in both a VOT series and a stop place series were correlated with acoustic-phonetic measurements on their own productions. This suggests that differences between individuals' perception are related to the idiosyncrasies of their production. These results fit well with theories of language acquisition emphasizing the role of individual experience. Since the voice that one has the most experience with and which one hears most often is one's own, an individual's own productions are likely to have an especially important role in the formation of perceptual expectations.

2:20

2pSC4. Variability in normal speech intelligibility: Separate and combined effects of talker-, listener-, and item-related characteristics. Ann R. Bradlow (Auditory Neurosci. Lab., Audiol. and Hearing Sci., Northwestern Univ., Evanston, IL 60208), Gina M. Torretta, and David B. Pisoni (Indiana Univ., Bloomington, IN 47405)

Various acoustic–phonetic and listener–transcription analyses were conducted using materials from two large databases of recorded speech. The first database consisted of sentence-length materials spoken by 20 talkers, along with intelligibility data in the form of transcriptions by ten native English listeners per talker. Acoustic–phonetic analyses of the digitized speech samples showed that talkers who exhibited a high degree of ''articulatory precision'' had higher intelligibility scores than talkers who produced more ''reduced'' speech. The second database consisted of a set of words spoken by ten talkers at three speaking rates, along with transcriptions by native English listeners. Analyses of the transcription accuracy scores showed a strong effect of inherent lexical characteristics, and a strong effect of speaking rate. Furthermore, the difficulties imposed by one factor, such as a fast speaking rate or an inherently difficult lexical item, could be overcome by the advantage gained through the listener's experience with the speech of a particular talker. Taken together, these data provide important information regarding the separate and combined effects of talker-, listener-, and item-related factors on normal speech intelligibility.

2:45–3:00 Break

3:00

2pSC5. Individual differences in the effects of stimulus variability and lexical difficulty in older adults. Mitchell S. Sommers (Dept. of Psych., Washington Univ., Campus Box 1125, St. Louis, MO 63130)

Sommers [M. S. Sommers, J. Acoust. Soc. Am. **101**, 2278–2288 (1997)] has recently demonstrated significant declines in both talker normalization and lexical discrimination in older listeners and suggested that these deficits contribute significantly to age-related impairments in speech perception. The present set of investigations were designed to: (1) determine whether individual differences in talker normalization and lexical discrimination are related to the speech perception abilities of older adults assessed under a variety of listening conditions; and (2) examine whether age changes in the two abilities are related to deficits in more general cognitive capacities. Results to date suggest that individual differences in talker normalization and lexical discrimination account for a significant percentage of the variance in speech perception scores among older listeners. In contrast, performance on tests of working memory, general intelligence, and selective attention were largely unrelated to either of the two abilities. The overall patterns of findings suggests that talker normalization and lexical discrimination are independent components of the speech perception system that exhibit considerable individual variation in the extent to which they are preserved in older adults. [Work supported by the Brookdale Foundation.]

3:25

2pSC6. Sensory and cognitive sources of individual differences in the ability to recognize speech. Charles S. Watson (Dept. of Speech and Hearing Sci., Indiana Univ., Bloomington, IN 47401)

Evidence in support of several generalizations about individual differences in speech recognition will be reviewed. First, the range of these abilities among normal-hearing adults appears to be reliable and large enough to be of some functional significance. Second, auditory spectral and temporal acuity, as measured with nonspeech sounds, fail to account for individual differences in speech recognition, although they may be weakly related to them. Third, a portion of the speech recognition variance among listeners is measurable with both speechreading (vision-only) and with auditory-only tasks, suggesting a central or cognitive origin. Fourth, the variety of types of distortion or transformation that fail to render speech unintelligible implies that very little necessary information is encoded in the fine structure of the speech waveform. Last, low correlations with general intelligence and with the ability to recognize environmental sounds suggest that speech recognition may be a relatively independent cognitive skill. Relevance of these conclusions to speech recognition by impaired listeners also will be considered. [Research supported by NIDCD and AFOSR.]

3:50

2pSC7. The scope of individual differences in cognitive models of spoken language understanding. Edward T. Auer, Jr. (Spoken Lang. Processes Lab., Human Commun. Sci. and Devices Dept., House Ear Inst., 2100 W. Third St., Los Angeles, CA 90057)

A model of spoken language understanding capable of accounting for data at the level of the individual should include: (1) a functional architecture and its processes; (2) specifications of how experience and development affect the architecture and its processes; and (3) knowledge about the range of variability that can be observed for the components of the system. Within this framework, individual differences are hypothesized to arise from the interaction of experience and biologically specified abilities. These individual differences result in systematic variation of performance in experimental tasks that can be identified with specific locations in the functional architecture and/or subprocesses. Taking into account variation due to individual differences, a model also should scale across normal and impaired populations. Evidence suggestive of the utility of this approach for understanding/predicting the performance of individuals will be presented from studies of speechreading, tactile-aid use, and cochlear-implant recipients. [This work was supported by NIH Grants DC-00695 and DC-02107.]

Session 2pSP

Signal Processing in Acoustics, Psychological and Physiological Acoustics and Speech Communication: Acoustics in Multimedia II: Head Related Transfer Functions

Isaac Graf, Chair
Speech and Hearing Sciences Program, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Invited Papers

1:30

2pSP1. Spatial perception, the acoustics of the external ear, and interactions with earphones. Edgar A. G. Shaw (Inst. for Microstruct. Sci., Natl. Res. Council, Ottawa, ON K1A 0R6, Canada)

In free-field listening, the perception of source elevation and discrimination between forward and rearward source positions appear to be largely dependent on the availability of high-frequency spectral cues associated with the mode structure of the concha and diffraction by the pinna, respectively. The role of the concha is particularly clear between 5 and 10 kHz, while pinna-diffraction effects are strongly developed between 2.5 and 6 kHz. When sounds are presented through earphones, the free-field characteristics of the ear are replaced by very different characteristics that are dependent on complex and unreliable interactions between the individual earphone and the individual ear. Since these interactions are highly dependent on earphone design, it may be appropriate to focus attention on the special characteristics that may be desirable in earphones intended for use in the creation of virtual acoustical environments. These may include coupling that (i) allows the response at the eardrum to follow the primary resonance curve of the individual ear and (ii) minimizes the inter- and intrasubject variations in response to other frequencies, including frequencies above 5 kHz where the wave characteristics of the concha come into play.

2:00

2pSP2. Application of head-related transfer functions for binaural synthesis. Dorte Hammershoei, Jesper Sandvad, and Henrik Moeller (Acoust. Lab., Aalborg Univ., Fredrik Bajers Vej 7 B4, DK-9220 Aalborg, Denmark, acoustics@kom.auc.dk)

Binaural synthesis offers a method for creating a truly immersive three-dimensional auditory environment. The sound output is produced by convolving a sound input with sets of head-related transfer functions (HRTFs). The result is a set of two signals typically reproduced by means of headphones, and the technique is therefore well suited for systems with limited hardware for sound reproduction, such as most multimedia systems. In the present paper, HRTFs are presented in the time and frequency domains, and are commented upon in relation to their application for binaural synthesis using digital signal processing. Results obtained from experiments conducted at the Acoustics Laboratory at Aalborg University in Denmark are reviewed, giving hints for the selection of filter lengths, update rate, and spatial resolution for dynamic environments.

Contributed Papers

2:30

2pSP3. Localization with individual and nonindividual binaural recordings. Henrik Moeller, Dorte Hammershoei, Clemen B. Jensen, and Michael F. Soerensen (Acoust. Lab., Aalborg Univ., Fredrik Bajers Vej 7 B4, DK-9220 Aalborg, Denmark, acoustics@kom.auc.dk)

The localization performance was studied when subjects listened (1) to a real sound field and (2) to binaural recordings of the same sound field, made (a) in their own ears, and (b) in the ears of other subjects. The sounds to be localized were loudspeaker reproductions of female speech at natural level, from 19 different positions in a standard listening room. The binaural recordings were reproduced by carefully equalized headphones. With individual recordings the performance was preserved compared to real life, whereas nonindividual recordings resulted in significantly more errors for sound sources in the median plane. Errors were seen in terms of confusion not only between nearby directions, but also between directions further away, such as between sound sources in front and behind the subject. However, nonindividual recordings made in the ears of a carefully selected ''typical'' subject resulted in a performance much closer to the real-life performance, although still inferior.

2:50

2pSP4. Localization with artificial head recordings. Jesper Sandvad, Flemming Christensen, Soeren K. Olesen, and Henrik Moeller (Acoust. Lab., Aalborg Univ., Fredrik Bajers Vej 7 B4, DK-9220 Aalborg, Denmark, acoustics@kom.auc.dk)

Twelve artificial heads, of which one was designed at the authors' laboratory, were evaluated in a localization test. The localization performance was compared in two situations: First, the subjects localized sound sources in a real sound field, then the localization test was repeated with artificial head recordings of the same sound field. The sounds to be localized were loudspeaker reproductions of female speech at natural level, from 19 different positions in a standard listening room. The artificial head recordings were reproduced by carefully equalized headphones. Results from eight of the heads showed an increased number of localization errors compared to real-life performance. The directions in the median plane were most often confused, not only with nearby directions, but also with directions further away. The number of errors was significantly higher than what can be obtained with recordings from a carefully selected human head. Results from the last four heads, including the one from the authors' laboratory, were not available at the time of abstract submission.

3:10

2pSP5. Multimicrophone arrangement as a substitute for dummy-head recording technique. Volker Mellert and Natalyia Tohtuyeva (Dept. of Phys., Univ. of Oldenburg, 26111 Oldenburg, Germany, mellert@aku.physik.unioldenburg.de)

Natural sound is usually recorded in head-related stereophony by ear microphones of a dummy head in order to ensure a virtual reality, e.g., for subjective sound-evaluation procedures. However, in many applications the use of a dummy head is disturbing. Additionally, it is difficult to change the geometry or shape of a dummy head in order to adjust a recording individually. The spatial information of the sound field, which is perceived in appropriate stereophonic reproduction, is encoded in the two signals of the ear microphones by recording the modifications of the wave field from the near-field diffraction of the obstacle ''dummy head.'' A diffraction pattern can be approximated by a sufficient number of monopole sources at appropriate locations, and adjusted in amplitudes and phases. This method of source substitution is used for an arrangement of pressure microphones. Each microphone signal is processed twice and added in the stereophonic channels. If locations and transfer functions are optimized, the two channels represent an approximation of the ear-microphone signals of the dummy-head's (or real head's) diffraction information. The preliminary results of the optimization and listening experiments are reported. [Work supported by the German Graduate School on Psychoacoustics.]

3:30

2pSP6. Measurement of the head-related transfer function at close range. Paul T. Calamia and Elmer L. Hixson (Dept. of Elec. and Comput. Eng., Univ. of Texas, Austin, TX 78712)

The head-related transfer function (HRTF) is generally considered to be a function of four variables: azimuth, elevation, range (all measured in a head-centered coordinate system), and frequency. The range dependence drops off with increasing distance, and thus HRTFs measured at a sufficiently large distance from the head may accurately represent the function at all points in the head's ''far field.'' This fact has resulted in a large body of research on far-field HRTFs, while very little has been done to characterize the function at close range. In this on-going work, HRTF measurements are being made on a KEMAR manikin at distances of 0.125, 0.25, 0.5, and 1.0 m. The experimental setup and procedure will be discussed, as will the results of the measurements. Preliminary analysis in anticipation of the authors' future work in close-range sound localization may also be presented.

3:50

2pSP7. Transforming the head-related transfer functions. Julio G. Farias (Dept. of Psych., California State Univ., 5151 University Dr., Los Angeles, CA 90032) and David R. Perrott (California State Univ., Los Angeles, CA 90032)

Auditory and visual localization of a real sound source under both natural and artificial listening conditions were examined. Five subjects were employed in a visual search task paradigm which contained some conditions that were aurally aided. The aurally aided conditions utilized white noise, emitted at the subjects horizon, from a single speaker at 72 dB(A). The paradigm also contained conditions in which a fully enclosed helmet was used to occlude the pinna in order to create the artificial listening conditions in which normal head-related transfer functions (HRTFs) are transformed. Reaction time and accuracy were simultaneously used within each condition to measure the extent to which normal HRTFs were transformed by the artificial listening conditions. The data were analyzed and significance was found between the natural and artificial conditions [$F(8,32)=3.59$; $p=0.005$]. There was an increase in both accuracy and reaction time between the natural and transformed HRTFs. (The experiment began with six subjects, but due to scheduling difficulties, their data were not available at the time this abstract was written. Their data will be included in the analysis for the presentation.)

TUESDAY AFTERNOON, 2 DECEMBER 1997 CALIFORNIA ROOM, 1:00 TO 4:45 P.M.

Session 2pUW

Underwater Acoustics and Engineering Acoustics: Advances in Underwater Range Technology, Underwater Communication and Sonars

Susan M. Jarvis, Chair
Naval Undersea Warfare Center Division, Code 382, Newport, Rhode Island 02841

Chair's Introduction—1:00

Invited Papers

1:05

2pUW1. An acoustic telemetry system for use on undersea ranges. Susan Jarvis and Fletcher Blackmon (Eng., Test and Eval. Dept., Naval Undersea Warfare Ctr. Div., Newport, RI 02841)

The underwater digital acoustic telemetry (UDAT) system is a set of underwater acoustic modems developed by the Engineering, Test and Evaluation Department at the Naval Undersea Warfare Center Division, Newport. The modems were designed to provide robust underwater acoustic communication at throughput data rates of approximately 900 and 1800 bits per second. Since February 1996 the modems have been used in nine exercises in five diverse deep-water and shallow-water environments. The results from the tests are very encouraging. Continuous, low-error transmission has been achieved over many minutes (~45 min). Detection ranges for the system have exceeded 4 km in shallow water and 8 km in deep water. In April and September 1997, the acoustic modems were used for continuous, full duplex, encrypted communications at the Atlantic Undersea Test and Evaluation Center (AUTEC). This paper will review the design of the UDAT system, discuss the results from the field tests, and present the current status of the modems. [Development of the modems was sponsored by Naval Air Systems Command, PMA-248.]

1:25

2pUW2. Ranges as undersea acoustic networks. Josko Catipovic (Naval Undersea Warfare Ctr., Newport, RI 02841)

Contemporary undersea acoustic ranges can be used as undersea local area networks to support cellular acoustic communication with mobile and fixed platforms and for associated local and remote sensing measurements. Bidirectional range nodes, particularly those containing multiple receiver hydrophones, inherently support high-speed acoustic communication with SSNs, UUVs, or other network users, and the range area of coverage can be used as an undersea cellular local area network. This talk will present an overview of acoustic local area network (ALAN) technology and highlight its compatibility with deployed and proposed ranges. A network communications protocol suitable for simultaneous communication with multiple platforms will be presented and shown to be compatible with the proposed East Coast Shallow Water Training Range. Experimental results in acoustic communication and networking relevant to range applications will be summarized. Adding an acoustic communication capability to a range greatly expands its usefulness. The talk will describe several applications ranging from SSN communications to cooperative UUV behaviors. Several related efforts are currently under development and ready for transition to an undersea range network. High-resolution signal processing required for acoustic communications also allows the use of range nodes for undersea remote sensing. Several applications of ranges to remote sensing of sound speed, current, and turbulence are realizable as extensions to processing required to support acoustic networking. An example network application for mapping current flows will be discussed, with extensions to travel-time tomography for sound speed inversion and scintillation tomography for volume mapping of turbulence.

Contributed Papers

1:45

2pUW3. Error-rate performance of convolutionally encoded, interleaved QPSK and 8PSK TCM underwater acoustic telemetry. Fletcher Blackmon (NUWC, 1176 Howell St., Newport, RI 02841)

A preliminary error-rate performance of comparison was conducted between two promising signal architectures for underwater acoustic telemetry modem use. The first candidate signal is a convolutionally encoded and interleaved QPSK modulation. The second candidate signal is an 8PSK trellis coded modulation (TCM). In both cases, the convolutional encoder is the same, i.e., $k=1$, $n=2$, and const$=7$. A short test in April of this year was performed at the Atlantic Undersea Test and Evaluation Center (AUTEC) in approximately 1000 m of water. Signals were transmitted from a seabottom mounted UQC transducer and were received by a hull-mounted transducer on a drifting surface vessel. One-hundred telemetry packets from each signal type were transmitted and were adaptively equalized within the receiver prior to decoding. The QPSK signal outperformed the TCM signal with bit error rates of 5e-6 and 1.4e-5, respectively.

2:00

2pUW4. Long-range signaling in shallow ocean transmission channels. Joseph A. Rice, Richard C. Shockley, Vincent K. McDonald (NRaD Code D881, San Diego, CA 92152), Dale Green (Torrey Sci. Corp.), and John G. Proakis (Northeastern Univ.)

Undersea acoustic signaling typically exploits the slowly varying, sparse impulse response of the physical channel. This paper addresses channels where the range is much greater than the water depth and where propagation ducts are absent. Here the existence of stable, discrete multipaths cannot be assumed. Transmission in these channels relies on forward-scattering from the rough seafloor and sea-surface boundaries. The Rayleigh and Ricean fading channel models are used to statistically describe the unpredictable transmission as a complex-valued Gaussian random process. Motion of the faceted scatterers at the time-variant sea surface further affects the transmitted signal by smearing the spectral content with Doppler spread. The prolonged signal duration and broadened signal bandwidth characterize a doubly spread transmission channel. An additional serious impediment is the dynamic and nonwhite channel noise. Coherent demodulation, channel-adaptive equalization, and matched-filter processing are inhibited in such doubly-spread, noisy channels. Robust 1- to 1000-bit/s undersea acoustic telemetry and ranging (telesonar) techniques are being developed by the U.S. Navy using digital coding, spread-spectrum modulation, diversity signaling, and noncoherent demodulation. Incidental benefits include CDMA networking, transmission security, and low power. [Work supported by ONR-321 and the SPAWAR SBIR Program.]

2:15

2pUW5. Probe signals for obtaining impulse response of doubly-spread channels. Vincent K. McDonald, Richard C. Shockley, Joseph A. Rice (NRaD Code D881, San Diego, CA 92152), and Dale Green (Torrey Sci. Corp.)

The U.S. Navy is developing undersea acoustic telemetry and ranging (telesonar) technology using shallow oceans as the physical transmission channel. The telesonar signaling philosophy is to achieve high reliability by designing for anticipated time and frequency spreading in the bounded, nonstationary, forward-scattered transmission channel. Obtaining the time-variant acoustic impulse response of representative channels supports appropriate signal design. These data allow measurement of time and bandwidth coherence, and aid understanding of boundary effects. Also, convolving signal waveforms with the impulse response simulates the influence of the transmission channel. This paper examines candidate probe signals for acquiring bandlimited acoustic impulse responses of shallow ocean channels. These probe signals are intended for systematic sampling of various littoral channels during telesonar sea tests. [Work supported by ONR-321.]

2:30

2pUW6. Spreading, coherence, fading, and noise in shallow-water channels. Richard C. Shockley, Joseph A. Rice, and Vincent K. McDonald (NRaD Code D881, San Diego, CA 92152)

During U.S. Navy undersea acoustic telemetry and ranging (telesonar) engineering tests, probe signals in the 8- to 16-kHz octave will be observed at various ranges and at various locales. Doppler spread and multipath spread will be measured for these waveforms as a function of time. In addition, simultaneous transmission of tonals will allow measurement of the correlation between fades at different frequencies, and hence the coherence bandwidth. Finally, ambient noise will be analyzed for insight into noise statistics and noise nonstationarity. The above information supports signal and receiver design by providing measured, as opposed to theoretical, channel characteristics. These data will also broaden the understanding of geographic and seasonal variabilty of undersea shallow-water channels. This paper describes procedures for extracting these channel statistics from recorded telesonar data. [Work supported by ONR-321.]

2:45

2pUW7. Noncoherent communication for adverse channels. Dale Green (Torrey Sci. Corp.) and Joseph A. Rice (NRaD Code D881, San Diego, CA 92152)

Undersea acoustic telemetry and ranging (telesonar) requires a realistic evaluation of the available physical transmission channel. Benign channels support high-bit-rate coherent transmissions, provided that a suitably powerful channel equalizer is available, and provided that covert signaling is

not required. Effective signaling in adverse channels requires accommodation of temporal and spectral dispersion and distortion, mitigation of interference, and asynchronous multiple access by many users. This is attained with frequency-hopped, M-ary FSK signals in which the elemental waveform ''chips'' are designed to tolerate channel spreading and nonstationarity. The M-ary nature of the signal reduces the needed SNR, while the frequency-hopping patterns inherently provide interference mitigation and multiple access. Specialized signal processing of the received signal effectively yields a channel with uniformly distributed noise in both time and frequency. Residual interference is removed by powerful nonbinary convolutional encoding and related sequential decoding. The sensitivity of signal acquisition and demodulation to errors in estimates of channel parameters is explored. Overly pessimistic estimates of spreading result in reduced channel access and increased interference at the processor output, while overly optimistic estimates result in signal loss and intersymbol interference. [Work supported by ONR-321 and NRaD telesonar Contract No. N66001-96-C-6003.]

3:00–3:15 Break

3:15

2pUW8. Adaptive beamforming for underwater acoustic communication in shallow water. Lester R. LeBlanc, Jochen R. Alleyne, and Pierre P. Beaujean (Dept. of Ocean Eng., Florida Atlantic Univ., Boca Raton, Fl 33431-0991)

An adaptive beamformer was developed for use with existing underwater acoustic modems and for measurement of the underwater acoustic communication channel. The output of the beamformer is made up of many independent channels, each derived from the application of orthogonal vectors to the array elements. Each channel represents a vertical beam that is focused in the direction of correlated energy arriving at the array of vertically spaced transducers. Within each beam, nulls are created in the direction of interference. Experiments were conducted in shallow water to characterize the acoustic communication channel at many frequencies and ranges so as to evaluate modem performance using various encoding methods and waveforms. It was found that high-reverberation levels exist at our test site in the shallow water near the shore environment. In this environment, adaptive processing reduced reverberation and transmission errors significantly. Wideband PSK was used to evaluate channel characteristics. In the process of testing the adaptive beamformer, waveforms were developed for a low-cost single-channel modem for use on AUVs. Results of the adaptive beamformer and the single channel modem are presented. [Work supported by ONR.]

3:30

2pUW9. The MPL sound velocimeter: An instrument for *in situ* sound velocity measurements in the deep ocean. Aaron D. Sweeney (Scripps Inst. of Oceanogr., Marine Physical Lab., 8602 La Jolla Shores Dr., La Jolla, CA 92093-0205, aaron@mpl.ucsd.edu), Fred N. Spiess, Dwight E. Boegeman, David M. Jabson, and Richard Zimmerman (Scripps Inst. of Oceanography, San Diego, CA)

A sound velocimeter designed to measure sound speed to a precision of 1 part in 10^5 has recently been tested at sea. The device was used in conjunction with a CTD to check agreement with sound speeds computed from popular empirical equations [C.-T. Chen and F. J. Millero, J. Acoust. Soc. Am. **62**, 1129–1135 (1977); V. A. DelGrosso, J. Acoust. Soc. Am. **56**, 1084–1091 (1974)]. Disagreements will be discussed. High-precision travel-time measurement is achieved through use of a phase-comparison technique. For this purpose, the device determines the number of cycles of the acoustic signal between the transmitted pulse and received echo, adjusting the frequency until the echo and transmitted pulse are in phase. The travel time is the product of the signal period and the number of cycles between the outgoing and incoming pulses. The velocimeter's capability of measuring travel time at several different frequencies near 4 MHz serves as a check on the timing consistency, since the travel time should be independent of frequency over this narrow range. The total distance traveled by the pulse is about 8.6 cm. With proper calibration, this pathlength can be known to a precision of 1 part in 10^5.

3:45

2pUW10. Quantitative comparison of two collocated seafloor hydrophone arrays in a shallow coastal water environment. J. Mark Stevenson (Acoust. Branch, Naval Command, Control and Ocean Surveillance Ctr., San Diego, CA 92152-6435, jms@nosc.mil), Steven L. Schmidt (NCCOSC RDT&E Div., San Diego, CA 92152), Gerald L. D'Spain, and William S. Hodgkiss (Scripps Inst. of Oceanogr., San Diego, CA 92152)

A novel, lightweight shallow-water hydrophone array was deployed and tested alongside a more conventional similar apparatus during the 1996 Adaptive Beach Monitoring Experiment (ABM-96). Controlled J-11 source tows (145-dB tones between 70 and 700 Hz) and local shipping provided a rich assortment of acoustic signals for intercomparison during a 6-h period. The new system, ULITE (ultralightweight sensor system), was deployed alongside the Marine Physical Laboratory's SRP Array in approximately 13-m water depth. The deployment of ULITE was conducted by hand from a 7-m-long vessel in approximately 1 h. The heaviest component of the system is a 5-kg anchor attached to the end of the ULITE array to prevent movement along the seafloor. Broadcast source tones were observed on single phones in the ULITE array. Received signal excess was, typically, 20–30 dB. The transmissions were simultaneously recorded on the nearby (within 100 m) SRP seafloor array. These data sets represent the first opportunity to conduct a quantitative intercomparison (i.e., single-phone and beamformed spectral ratios) between the ULITE array and a time-proven conventional array. Since the individual hydrophones of both systems are calibrated, a meaningful comparison of absolute levels is possible. [Work supported by ONR 321SS.]

4:00

2pUW11. A forward-scan sonar system for small unmanned submersibles. Joe Cuschieri and Lester LeBlanc (Dept. of Ocean Eng., Florida Atlantic Univ., Boca Raton, FL 33431)

Over the past 3 years, work has been devoted to the development of a forward-looking sonar system that can be used in small unmanned submersibles. One of the sonar systems developed, a forward scan (FS) sonar, generates side-scan-type images with the exception that multiple scan lines (120 lines in a ± 60-deg forward-looking sector) are simultaneously formed, using digital processing of the signals received by the 64 elements of the receiver hydrophone array. The system operates at 230 kHz with a 10% bandwidth using a frequency chirp. The signals received by each element of the receiver hydrophone are sampled and digitized. Since the signals only have 10% bandwidth, the digital data are filtered and decimated to reduce the volume of data processed. Test tank and at-sea sonar images have been obtained using the FS sonar at the rate of one image per second using two skybolt quad i860 processors. The target is to generate four images per second. [Work sponsored by ONR.]

4:15

2pUW12. Three-dimensional forward look sonar system. Joe M. Cuschieri and Lester LeBlanc (Dept. of Ocean Eng., Florida Atlantic Univ., Boca Raton, FL 33431)

As part of an ongoing research effort related to new sonar systems, a system that can directly generate 3-D images of the underwater terrain has been developed. This 3-D forward-look (3-D FL) sonar consists of a separate projector and receiver hydrophone. The projector has a 1 deg fan beam, which is electronically scanned in 1-deg steps within the forward sector of view (± 60 deg). The receiver hydrophone array oriented crosswise to the projector array can simultaneously form 100 beams within ± 50 deg, generating a profile for each of the positions of the projector fan beams. The sequence of profiles thus generated constitute a 3-D image of the underwater terrain or bottom targets. The system operates at 230 kHz with a 20% bandwidth. To speed the generation rate of the profiles a

pseudorandom signal is used. The level of data processing required to generate the profiles is, however, significant. Test tank and stationary at-sea image data have been collected and will be presented. [Work sponsored by ONR.]

4:30

2pUW13. Recent results from the DARPA synthetic aperture sonar. Matthew A. Nelson (Dynam. Technol., Inc., 21311 Hawthorne Blvd., Ste. 300, Torrance, CA 90503, mnelson@dynatec.com)

A synthetic aperture sonar (SAS) system has been developed by Dynamics Technology, Inc. (processing lead), and Hughes Aircraft Naval and Maritime Systems (hardware and operation). The system operates at 50 kHz and features a 3.2-m-long, 16-element linear array and a data acquisition system that records phase and amplitude information for each element. Towbody motion is estimated from information contained in the sonar data through the use of a type of long-track interferometry, aided by data from a simple motion compensation suite consisting of Doppler sonar and a single gyro. This combination compensates motion to levels correctable by standard SAR autofocus techniques modified for SAS. The system has demonstrated, at long ranges (up to 1000 m), high-resolution images (better than 20 cm), previously achievable only with high-frequency sonars at close range. The low-carrier frequency allows high-resolution imaging of the interior structures of submerged objects, recently demonstrated by SAS imagery revealing the rib and spar structure inside a wrecked PB4Y-2 airplane in Lake Washington. [Work supported by DARPA.]

Session 3aABa

Animal Bioacoustics: C. Scott Johnson Session: Whale and Dolphin Acoustics II

David A. Helweg, Chair
Code D-351, RDT&E Division, NCCOSC, 49650 Acoustic Road, Room 108, San Diego, California 92152-6254

Chair's Introduction—8:00

Invited Paper

8:05

3aABa1. The Office of Naval Research Program to determine the effects of man-made sound on marine mammals. Robert C. Gisiner (Office of Naval Res., Code 335, 800 North Quincy St., Arlington, VA 22217-5660)

Over the past 5 years the Office of Naval Research (ONR) has maintained a coordinated interdisciplinary research program focused on the auditory capabilities of marine mammals, and the potential impacts man-made noise might have by damaging or obscuring that sensory capability. Auditory studies have included behavioral determinations of thresholds to low-frequency sound, effects of depth (pressure) on auditory function, and temporary threshold shift studies. Anatomical data, auditory-evoked potentials and otoacoustic emissions have been explored with the goal of providing larger sampling of populations and enabling sampling of the larger and rarely encountered species, such as baleen whales. The data collected thus far have further strengthened the view that marine mammal hearing represents the pinnacle of development of auditory sensing. With our own use of underwater acoustic sensing in its infancy, it would be wise to closely monitor those activities which might obscure our window into the ocean environment as well as that of its inhabitants who have led us this far.

Contributed Papers

8:25

3aABa2. Acoustic and visual studies of blue whales near the California Channel Islands. Arthur M. Teranishi, John A. Hildebrand, Mark A. McDonald, Sue E. Moore (Scripps Inst. of Oceanogr., Marine Physical Lab., Univ. of California, San Diego, CA 92093), and Kate Stafford (Natl. Oceanic and Atmospheric Administration, Newport, OR 97365)

Preliminary results are presented from a combined acoustic and visual study of blue whales in the Channel Islands National Marine Sanctuary off the California coast, conducted during four 3-day expeditions in July 1996. Acoustic recordings were collected using GPS positioned sonobouys arrays, with up to five sonobuoys deployed at approximately 1-km spacing. Visual observations were made to estimate the number of blue whales in the vicinity of the arrays, and to record whale behaviors. Significant numbers of blue whales (5–15) were sighted each day, and as many as 100 blue whale calls per hour were recorded. About one-third of the time, large ships transited through this area, providing loud low-frequency sound sources to examine blue whale sensitivity to noise. Low-frequency blue whale calls, similar to those previously designated ''A'' and ''B'' in the literature, were recorded. Additional blue whale calls include an approximately 4-s duration frequency downswept tone between 30 and 80 Hz, often heard at about the same intensity as the more typical ''A'' and ''B'' calls. Cross correlation of signals received by sonobuoys provides information to localize calling whales. The location accuracy relative to the different call types will be discussed.

8:40

3aABa3. Detections of singing humpback whales (*Megaptera novaeangliae*) across the northeast Pacific during the SWAPS97 sperm whale cruise. Thomas F. Norris (Bio-waves, 8455 Kingsland Rd., San Diego, CA 92123, tomnorris1@juno.com), Jay Barlow (NOAA Southwest Fisheries Sci. Ctr., La Jolla, CA 92038), and Mark McDonald (Laramie, WY 82070)

The NOAA Southwest Fisheries Science Center conducted a 3-month (8 March–9 June 1997) Sperm Whale Abundance and Population Structure (SWAPS97) cruise to visually and acoustically survey sperm whales in the northeast Pacific. A custom-designed towed hydrophone array system was deployed to detect and determine bearings to sperm whale ''clicks.'' However, this system also proved effective in detecting and recording songs of humpback whales. Additionally, sonobuoys were deployed regularly throughout the cruise to monitor and record baleen whales' sounds. The timing and location of these surveys corresponded well with the northbound migration of humpback whales from known tropical breeding areas to northern latitude feeding areas and resulted in numerous (~130) detections of singing humpback whales. Most of these detections occurred in pelagic waters (i.e., away from breeding areas), indicating that humpback whales sing during their northward migration. Of greater interest were clusters of detections 400–600 miles off the coast of California. It is possible these detections correspond to animals migrating north from the Revillagigedo Islands, Mexico. Data on singing humpback whales collected during the SWAPS97 cruise will be summarized, and implications regarding the biology and management of humpback whale populations will be discussed. [Work supported by NOAA/NMFS and ONR.]

8:55

3aABa4. New vocalization definitively linked to the minke whale. Jason Gedamke, Daniel P. Costa (Inst. of Mar. Sci., A316 EMS, UCSC, Santa Cruz, CA 95064, jgedamke@cats.ucsc.edu), and Andy Dunstan (R/V Undersea Explorer, Port Douglas, Queensland, Australia)

The description of minke whale (*Balaenoptera acutorostrata*) vocalizations is sparse with little recent study. In addition, the minke and some of its assumed vocalizations are not definitively linked. To address this problem, recordings were made at a unique site off the Great Barrier Reef where minkes maintain long, close contacts (to 10 h) with vessels. During these encounters a 200-m, four-element hydrophone array (Hi-Tech HTI-SSQ41B, 10 Hz–30 kHz) was deployed and digitally recorded onto a Tascam DA-38 (0–22 kHz). Time of arrival differences between hydrophones were used to obtain hyperbolic fixes on nearby sound sources. With this method, a new vocalization consisting of three short pulses (pulse duration 100 ms, peak energy 200–250 Hz, with significant energy between 1.5 and 2 kHz) followed by a slight upsweep (trail of 1 s) has

been definitively linked to the minke. Source levels of at least 155 dB *re*: 1 μPa were estimated. Similar vocalizations with varying pulse numbers (to at least 7), and low frequency grunts and moans (80–4000 Hz) were recorded. Approximately 19 h were recorded in the close presence of 45 minke whales along with 18 h of ambient recordings. These recordings greatly contribute to the sparse database of known minke vocalizations. [Work supported by ONR and the R/V Undersea Explorer.]

9:10

3aABa5. Low-frequency whale calls recorded on hydrophones moored in the eastern tropical Pacific. Kathleen M. Stafford (Oregon State Univ., 2030 S. Marine Sci. Dr., Newport, OR 97365) and Christopher G. Fox (NOAA/Pacific Marine Environ. Lab., Newport, OR 97365)

Six moored hydrophones have been recording low-frequency sound in the eastern tropical Pacific (ETP) since May, 1996 (sample rate 100 Hz, low-passed at 40 Hz). These hydrophones are spaced widely: they are moored at latitudes 8 N, 0 and 8 S along longitudes 95 W and 110 W. These distances make it very unlikely that two or more hydrophones will simultaneously record sound from one animal. Continuous spectrograms of data from every third day for each hydrophone were examined for whale calls. Calls attributed to eastern north Pacific blue whales (*Balaenoptera musculus* spp.) were the most commonly recorded recognizable calls. Additionally, western north Pacific blue whale calls were detected as were calls that resemble those recorded in the presence of blue whales off Chile. Pulse series similar to those which have been attributed to minke whales (*B. acutorostrata*) were also recorded. A number of unidentified ''biologicals'' were recorded. Among these are groups of 1–6 low-frequency pulses with regular interpulse spacing. The data from the eastern hydrophones show more call variability and abundance than do those to the west. This array of hydrophones is semipermanent and should provide years of acoustic data from the ETP.

9:25

3aABa6. Broadband characteristics of spinner dolphin (*Stenella longirostris*) social acoustic signals. Marc O. Lammers, Whitlow W. L. Au, and Roland Aubauer (Marine Mammal Res. Program, Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734, lammers@hawaii.edu)

A predisposition generally exists among delphinids for hearing high-frequency signals, raising the question: How much do dolphins actually use the ultrasonic range for social signaling? To answer this question, a low-budget, portable broadband recording system was developed to collect whistles and burst pulses from a population of Hawaiian spinner dolphins resident off the leeward coast of Oahu, Hawaii. This system employs a PCMCIA A/D converter to digitize signals directly into a laptop computer and has a maximum upper frequency limit of 140 kHz. Signals collected using this system reveal that spinner dolphins use the ultrasonic range extensively. Although the fundamental frequencies of whistles tend to be restricted below approximately 23 kHz, strong harmonics are often clearly visible up to 70 kHz. Even more high frequency in nature are the burst pulses. These signals often have little or no energy below 20 kHz and can extend beyond the recording limits of the system. Furthermore, high-frequency burst pulses are pervasive in the acoustic output of spinner dolphins, making it reasonable to assume that the ultrasonic range may, in fact, be important not only for echolocation, but also social signaling purposes.

9:40

3aABa7. The neutral network analysis of false killer whale (*Pseudorca crassidens*) vocalizations. Scott O. Murray (Brain Development Lab., Psych. Dept. 1227, Univ. of Oregon, Eugene, OR 97403), Eduardo MercadoIII, and Herbert L. Roitblat (Univ. of Hawaii, Honolulu, HI 96822)

This study reports the use of unsupervised, self-organizing neutral networks to categorize the repertoire of false killer whale vocalizations. Self-organizing networks are capable of detecting patterns in their input and partitioning those patterns into categories without requiring that the number or types of categories be predefined. The inputs for the neural networks were two-dimensional characterizations of false killer whale localizations where each vocalization was characterized by modulations in duty cycle and peak frequency [Murray *et al.* (in prep.)]. The first neural network used competitive learning, where neurons in a competitive layer distributed themselves to recognize frequently presented input vectors. This network resulted in classes representing typical patterns in the vocalizations. The second network was a Kohonen feature map which organized the outputs topologically, providing a graphical organization of pattern relationships. The networks performed well as measured by (1) the average correlation between the input vectors and the weight vectors for each category, and (2) the ability of the networks to classify novel localizations. The techniques used in this study could easily be applied to other species and facilitate the development of objective, comprehensive repertoire models.

Session 3aABb

Animal Bioacoustics: Biologically Inspired Acoustics Models and Systems I

V. Ashley Walker, Chair
Department of Artificial Intelligence, University of Edinburgh, 5 Forrest Hill Road, Edinburgh EH1 2QL, Scotland

Chair's Introduction—10:00

Contributed Papers

10:05

3aABb1. Bionic synthetic aperture sonar. Richard A. Altes (Chirp Corp., 8248 Sugarman Dr., La Jolla, CA 92037, altes@msn.com) and Patrick W. B. Moore (NRaD D351, San Diego, CA 92152-6254)

Dolphinlike sonar signals are used to obtain synthetic aperture images of complicated, three-dimensional (3-D) objects from transmitter–receiver positions that are above the plane of effective target rotation. This geometry allows 3-D information to be incorporated into a SAS image. The 3-D information can be represented as a set of spatially registered 2-D (x,y) pixel maps, where each map corresponds to a different focal point in elevation (z). Similar spatially registered environmental maps are used for sensor fusion in the mammalian superior colliculus and in the optic tectum of reptiles and fish. The constant-z maps can be combined to improve the 2-D representation of a 3-D object. For example, the image at a given x,y position can be focused by using pixels from the constant-z map that maximize a local focus criterion at position x,y in the map. Alternatively, a surface recognition criterion yields the z coordinate of the surface of a 3-D object at a given x,y point. The resulting 3-D representation of the object's surface corresponds to a visionlike representation. These representations can be combined with other spatially registered maps from the visual domain for bionic sensor fusion.

10:20

3aABb2. Aspect-independent classification of ''dolphin'' ensonified mines using Choi–Williams representations. David A. Helweg and Patrick W. B. Moore (NCCOSC RDTE DIV., Code D351, 49650 Acoust. Rd., Rm. 108, San Diego, CA 92152-6254)

Contemporary anti-invasion mines have aspect-dependent shapes, making discrimination of mines from nonthreat objects, and classification of mines, a difficult task for human mine countermeasures personnel. Shallow-water (SW) noise and bottom reverberation substantially degrade the acoustic structure of mine echoes. Traditional MCM target strength and FFT classifiers are not effective under these conditions. The testing of a novel neural network classifier to solve the task of classifying mines in the SW acoustic environment was begun. Three mine types were ensonified at 1 deg rotations using synthetic dolphin clicks. A learning vector quantization network was trained to classify the mines using the Choi–Williams joint time-frequency distributions of echoes. A training set of 36 echoes per mine (5, 15, 25, . . . , 355 deg) was created, with the remaining echoes (0, 10, 20, . . . , 350 deg) reserved for generalization testing. The network correctly classified the mines at novel orientations with 85% accuracy. Performance using CWD will be contrasted with biomimetic sonar target classifiers under varying signal-to-noise ratios. [Research funded by ONR Underseas Active Signal Processing Program.]

10:35

3aABb3. Visualizing dolphin sonar signal generation using high-speed video endoscopy. Ted W. Cranford (QMC and Biology Dept., San Diego State Univ., 5500 Campanile Dr., San Diego, CA 92182-4614), William G. Van Bonn (Upstream Assoc., San Diego, CA 92166), Monica S. Chaplin, Jennifer A. Carr, Trish A. Kamolnick (SAIC, San Diego, CA 92121), Donald A. Carder, and Sam H. Ridgway (NCCOSC RDTE DIV., San Diego, CA 92152)

Determining the site of the delphinid sonar signal generation has eluded cetologists for decades. Activities within the pharyngeal and nasal cavities of two bottlenose dolphins (*Tursiops truncatus*) were studied during sound production and echolocation. A high-speed dual-camera video system provided synchronized windows for recording two concomitant events: (1) movements visible through an endoscope and (2) oscilloscope traces of acoustic pressure at a hydrophone placed near the animal's head. Dolphins have two tissue complexes, one located on either side, and just above, the membranous nasal septum [Cranford *et al.*, J. Morph. **228**, 223–285 (1996)]. They apparently generate acoustic pulses by pushing air across sets of internal ''lips.'' The acoustic pulse occurs coincident with one oscillatory cycle of the lips. Changes in the acoustic pulse repetition rate and the vibration cycles of the lips are simultaneous, indicating that their rates and periods are synchronous. No other structures were found to vibrate in synchrony with each acoustic pulse generation event. The palatopharyngeal muscle complex compresses air for the system. These observations settle a long-standing controversy over the site of biosonar signal generation in odontocetes and open a vista of potential avenues for future investigations. [Work supported by the Office of Naval Research.]

10:50

3aABb4. Biosonar dynamics of the bottlenose dolphin in VSW search and detection tasks. John E. Sigurdson (NRaD, 53405 Front St., Rm. 200, San Diego, CA 92152)

The components of the dolphin's biosonar are well studied, but their combination and dynamic adjustment during echolocation in open waters are poorly understood. A lack of adequate measurement and research techniques has delayed progress, but new technology and methods now allow for the continuous recording of the animal's interpulse-interval, head-attitude, azimuth, and pulse waveform during search and detection in very shallow open water. Such measurements were made in controlled tests with bottom objects in random positions on random sand backgrounds. The data provided a 3-D description of the variation in biosonar search behavior as a function of object position and practice effects. Basic findings included the adaptation of the search pattern to the spatial distribution of previously detected objects, correlations between the animal's orientation, pulse rate, and object location, as well as an inverse relation between the number of pulses on object and the object's distance. Steady pulse rates on nearby objects were also observed with only half of the ''over-two-way'' interval previously reported. Coincident with these dynamic adjustments, the pulse spectra displayed considerable variation with detected objects, changing distance and grazing angles. The spectral data also indicate a dual-frequency source.

11:05

3aABb5. Modeling of penetration of dolphin sonar clicks into ocean sediments. Nicholas P. Chotiros (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

The penetration of dolphin sonar clicks into the ocean sediment is modeled using Biot's theory of sound propagation in porous media. A range of sediment types are considered, from soft mud to sand. The expected signal loss as a function of depth and grazing angle are computed. At any given value of loss, the penetration may vary by more than an order of magnitude depending on sediment type. The penetration is modified by critical angle effects and interference between different sediment acoustic waves. In modeling the acoustic scattering from buried objects, it is not sufficient to compute the net acoustic pressure, but also the pressures of each contributing sediment acoustic wave component. [Work supported by ONR code 342.]

11:20

3aABb6. Classification of dolphin echolocation clicks by means of energy and frequency distributions. Dorian S. Houser (Dept. of Biol., Earth and Marine Sci., A316, Univ. of California, Santa Cruz, CA 95064, houser@biology.ucsc.edu), David A. Helweg, and Patrick W. Moore (NCCOSC RDTE DIV. D351, San Diego, CA 92152-6254)

Echolocation clicks ($N \sim 30\,000$) were collected from an Atlantic bottlenose dolphin (*Tursiops truncatus*) performing object discrimination tasks. Eight categories of clicks were identified using the spectral conformation and relative position of -3- and -10-dB peaks. A counterpropagation network utilizing 16 inputs, 50 hidden units, and 8 output units, was trained to classify clicks using the same spectral variables. The network classified novel clicks with 92% success. Additional echolocation clicks ($N > 24\,000$) from two other dolphins were submitted to the network for classification. Classified echolocation clicks were analyzed for animal-specific differences, changes in predominant click type within click trains, and task-related specificity. Differences in animal and task performance may influence click type and click-train length.

11:35

3aABb7. Seismic transmission of elephant vocalizations and movement. Caitlin E. O'Connell (Ctr. for Animals in Society, School of Veterinary Medicine, Univ. of California, Davis, CA 95616), Byron T. Arnason (Tezar, Inc., Austin, TX 78731), and Lynette A. Hart (Univ. of California, Davis, CA 95616)

Seismic signals were recorded during periods of vocalizations and movement by African and Asian elephants. Initially, vocalizations were detected in both microphone and geophone recordings. To test whether these sounds are propagated separately in the air and the ground, recordings were made with a four-channel data acquisition system, processing signals simultaneously from microphone/geophone pairs at near and remote points ranging from 20–130 m distances from the elephants. These recordings indicate that elephant vocalizations and movement produce Rayleigh waves, a type of ground surface wave. This paper quantifies the propagation of elephant vocalizations in the air and ground and compares the rates of signal attenuation within each of the two media. A normalized cross-correlation coefficient was performed on rumble and movement signals in the ground, resulting in a time lag between near and remote sensors corresponding to Rayleigh wave velocity. Rayleigh waves travel slower than airborne sound waves (approximately 250 m per second versus 340 m per second), but they can travel further and more reliably than those transmitted through the air. The presence of seismic properties in elephant vocalizations and movement suggests a possibility for their use in long distance communication.

11:50

3aABb8. Surface meteorology and animal communication. David Larom (Dept. of Biol., Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92138), Michael Garstang (Univ. of Virginia, Charlottesville, VA 22903), Katharine Payne (Cornell Univ., Ithaca, NY 14850), and Richard Raspet (Univ. of Mississippi, Oxford, MS 38677)

Vertical temperature and wind gradients exert powerful and predictable influences on the area ensonified by animal vocalizations. Computer modeling of 15- and 30-Hz sound propagation in measured atmospheric conditions predicts that infrasonic African elephant calls can have ranges exceeding 10 km and will be highly directional in the presence of wind shear. The calling area is maximized under temperature inversions with low wind speeds. The calling area can expand and contract by an order of magnitude in 24 h. Topography, regional weather patterns, seasons, and climate variation modify this cycle. Similar influences affect the somewhat higher frequency calls of lions and may be a selective pressure toward their crepuscular and nocturnal calling behavior. Coyotes and wolves, which also live in areas with strong and prevalent nocturnal temperature inversions, show similar calling patterns, maximizing their chances of being heard over the longest possible distances. The pronounced dawn and evening vocalization peaks in other animals including birds, frogs, and insects may reflect the same influences, in combination with other factors which selectively limit high-frequency sound propagation. Atmospheric conditions therefore need to be taken into account in many studies of animal behavior. [Work supported by NSF Grant Nos. ATM92-07924 and ATM-9529315. Thanks to Peter Marler for useful comments.]

Session 3aEA

Engineering Acoustics: Fiber Optic Sensors for Acoustic Applications

Joseph A. Bucaro, Chair
Naval Research Laboratory, Code 7130, 4555 Overlook Avenue, S.W., Washington, DC 20375-0002

Chair's Introduction—9:30

Invited Papers

9:35

3aEA1. Fiber-optic sensors for acoustic array applications. Anthony Dandridge (Naval Res. Lab., Code 5670.1, 4555 Overlook Ave., S.W., Washington, DC 20375-5320, adandrid@ccf.nrl.navy.mil)

Over the past 7 years there have been a number of impressive demonstrations of fiber-optic acoustic sensor systems with surveillance grade capabilities. These demonstrations range from single-element and small six-element arrays tests conducted in the Arctic (1990 and 1992), to a 128-channel capacity (64-populated channels), horizontal line array deployed off Point Loma in 1996. These fixed, surveillance arrays have wider bandwidths and larger dynamic ranges that their piezoceramic counterparts. As well as fixed arrays, hull-mounted array demonstrations (such as the lightweight wide aperture array) have met the stringent Navy performance specifications while eliminating all the outboard electronics to reduce the systems life-cycle costs. The lightweight planar array approach is a viable candidate for the NSSN WAA. To reduce system costs further, planar array approaches using fiber optics appear to be the candidate of choice for the next generation of submarine mounted arrays. The other area attracting attention is that of the all-optical towed array, where the reduction in wet-end cost is the prime driver. In this system concept, engineering functions required for the operation of the array are accomplished by a suite of fiber-optic sensors for temperature, pressure, heading, roll, pitch, and strain.

10:00

3aEA2. Fiber-optic sensors for underwater sonar applications. Marilyn J. Berliner (Code 2141, Naval Undersea Warfare Div., 1176 Howell St., Newport, RI 02841-1708, berliner@coma.npt.nuwc.navy.mil)

An overview of fiber-optic sensor designs and performance for underwater sonar applications will be discussed. The sensor designs range from simple, coated fibers to mandrels wrapped with single or multiple layers of fiber. The measurable response of fiber-optic sensors is through changes in the phase of light in the fiber due to the internal components of strain. The assumed strain in models of the acoustic response of optical fiber sensors differs for the various sensor designs and often are derived from an analysis of deformations of a complex, composite structure. Another consideration in the design of fiber-optic sensors is structural resonances, which may interfere with acoustic performance. Comparisons of modeled and measured acoustic and static pressure sensitivities of various fiber-optic sensor designs will be presented along with laboratory measurements of vibration response.

10:25

3aEA3. Fiber-optic sensors for structural–acoustic applications. B. H. Houston and J. A. Bucaro (Naval Res. Lab., Code 7130, 4555 Overlook Ave., S.W., Washington, DC 20375-5350, jbucaro@ccf.nrl.navy.mil)

Fiber-optic sensors under development for structural-acoustics applications will be discussed. These applications include the structural-acoustic monitoring and control of submarine and air-frame systems. In both cases, the ''structure'' is more or less cylindrical in shape, and the goal is to apply sensors to the structure, ideally, acoustic pressure and velocity sensors, but in some cases strain sensors, so as to be able to ''know'' through the application of suitable algorithms the exterior sound field (in the submarine case) or the interior noise field (for the air-frame system). In addition to the details of the particular fiber sensors, which have been successfully developed, a view will be presented as to what new and improved sensors are required for these applications. [Work supported by the Office of Naval Research.]

10:50

3aEA4. Fiber-optic acoustic and vibration sensors and systems. David A. Brown (Dept. of Elec. and Comp. Eng., Ctr. for Marine Sci. and Technol., Univ. of Massachusetts at Dartmouth, Dartmouth, MA 02747 and Naval Undersea Warfare Ctr., Transduction Group, Code 2131, Newport, RI)

Fiber-optic interferometric acoustic sensors and accelerometers have been developed for a wide variety of marine and geophysical applications. They offer advantages including compatibility with optical fiber telemetry, light weight, and removal of electronic in the wet-end. This presentation will review sensor designs used in sonar applications. Interrogation and multiplexing schemes using 3×3 couplers and fiber-optic Bragg gratings for acoustic and vibration applications will also be presented. [Work supported by ONR and NUWC.]

Session 3aED

Education in Acoustics: Hands-on Acoustics for High School Students

Scott D. Sommerfeldt, Cochair
Department of Physics, Brigham Young University, 277 Fletcher Building, Provo, Utah 84602

Uwe J. Hansen, Cochair
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 San Diego area. Presenters will introduce each topic with the associated instrumentation in 3–5 minutes. 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, among others, will be acoustic levitation, active noise control, spectral analysis of musical sounds, observation of normal mode vibrations by observing Chladni patterns and by probing the near field with a microphone. 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.

Contributed Paper

10:05

3aED1. Research projects in acoustics for high school mentorship students. Murray S. Korman (Dept. of Phys., U.S. Naval Acad., Annapolis, MD 21402) and Bonnie S. Dixon (Anne Arundel County Public Schools, Annapolis, MD 21401)

Since 1983, the Anne Arundel County Public Schools' Gifted/Talented/Advanced Program Office and the Naval Academy staff have developed a partnership in a program that is commonly called ''Mentorship at the U.S. Naval Academy.'' The program joins two special needs: the students need for advanced academic experiences and research projects along with the Naval Academy's (an undergraduate school) need for research assistants to help the professors. The mentorship program is an intentional nurturing process for high school students working in an academic, scientific, and authentic research project environment. These projects require skills that are beyond what the classroom teachers and the curriculum can offer to advanced students. The process involves an optimal match between student and mentor—involving applications, interviews, selection, project involvement, and growth. Students are expected to work 132 h over a full year, receiving 1/2 a Carnegie credit unit for every 66 h worked—that counts as additional credit only. The Physics Dept. ''mentees'' do research in acoustics. Over the years, seniors have performed experiments involving sound generation by turbulent jets, scattering of sound by turbulence or by bubbly flow, nonlinear acoustics, chaos, Fourier analysis, electronic or computer projects, music acoustics, and sonar projects like ray tracing.

10:20–12:00 noon

Demonstrations and Hands-on Participation

Session 3aNS

Noise: Allocation of Hearing Loss

Larry H. Royster, Chair
MAE Department, North Carolina State University, Raleigh, North Carolina 27695-7910

Chair's Introduction—8:30

Invited Papers

8:35

3aNS1. Allocation of hearing loss in civil litigation. Fred Krutz (Forman, Perry, Watkins, Krutz & Tardy, 188 E. Capitol St., Ste. 1200, Jackson, MS 39201, fred@fpwk.com)

The defense of product liability cases filed by plaintiffs alleging that their hearing loss was caused by defendants' noise-making products always involves the analysis of all relevant evidence to determine the relative contributions of different causes (occupational noise, recreational noise, aging, etc.). That is because in most jurisdictions the plaintiff will be required to prove that exposure to the defendant's product was a ''substantial factor'' in causing his or her hearing loss, and because the size of any verdict may depend on the magnitude of that contribution. Moreover, for that same reason and also in order to determine whether a statute of limitations defense is available in the case, it is often important to estimate the progression of a plaintiff's hearing loss over time. All relevant evidence, including quantitative methods based on epidemiological standards supported by expert testimony, should be used to assist the jury in determining whether the plaintiff has proven his or her case by a preponderance of the evidence (50.1% vs 49.9%).

8:45

3aNS2. Apportionment of hearing impairment causation—Are we there yet? David M. Lipscomb (Correct Service, Inc., P.O. Box 1680, Stanwood, WA 98292)

The ability to discern between etiological factors in hearing impairment depends upon using an accurate and fair method for apportionment. The popular method for division of causation assumes: (1) that pure tone audiometric information reflects the condition of the hearing mechanism (validity); and (2) that the AMA handicap percentage calculations accurately represent the degree of handicap experienced by a given individual (validity, again). Neither of these assumptions can be substantiated. Histological evidence contradicts the notion that pure tone hearing testing gives adequate information on the true status and condition of inner ear sensory neuroepithelium for purposes of apportionment. To ''break apart'' (allocate) complex causation in audiometric data, the effects of etiologies must be ''additive'' or they must combine in some measurable fashion. There is no convincing evidence regarding the additive nature of two or more etiologies with respect to pure tone audiometric data. In fact, hair cell injury and destruction is due to complex interaction between etiological factors.

8:55

3aNS3. Can one allocate noise-induced hearing loss retrospectively for individuals? William W. Clark (Central Inst. for the Deaf, 818 S. Euclid, St. Louis, MO 63110, clark@cidmac.wustl.edu)

For decades, physicians and other clinicians have been asked by the legal profession to answer two simple questions about individuals presenting with marked or measurable hearing loss: (1) How much of the observed hearing loss was caused or contributed to by excessive exposure to occupational noise? (2) How did this hearing loss accumulate over time? Until recently, answers provided represented no more than clinical guesswork. However, with the adoption of the international standard ISO 1999, ''Acoustics-Determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment,'' and its heterozygous twin, ANSI Standard S3.44, of the same title, clinicians and scientists now have a tool, or framework from which they may construct answers to the two questions. The precision of the answers, however, depends entirely upon the precision of the framework and data underlying the standard. This presentation will summarize the weaknesses in the ISO standard and will support the conclusion that one still cannot, with any reasonable degree of scientific certainty, answer either question for an individual.

9:05

3aNS4. Allocation of hearing loss. Robert A. Dobie (Dept. of Otolaryngol.—Head and Neck Surgery, Univ. of Texas Health Sci. Ctr., 7703 Floyd Curl Dr., San Antonio, TX 78284, dobie@uthscsa.edu)

In medical–legal evaluation of persons with hearing loss, *allocation* means the clinical process of estimating the relative contributions of different causes (noise, aging, etc.) or of different time periods. Allocation is not age correction; the practice of age correcting an audiogram prior to calculation of a hearing handicap score is fundamentally unfair. Allocation is an exercise of informed clinical judgment, requiring a solid understanding of the epidemiology of age-related and noise-induced hearing loss. While data specific to the individual, such as history, noise exposure measurements, and serial audiograms, should be given the greatest weight, quantitative methods based on epidemiological standards can also assist the clinician in making reasonable and unbiased allocation estimates.

9:15

3aNS5. Making decisions about populations' and individuals' hearing thresholds in terms of their life-long noise exposure histories. Larry H. Royster (North Carolina State Univ., P.O. Box 7910, Raleigh, NC 27695-7910, effective_hcps@compuserve.com)

Industrial safety and health personnel and professionals with various backgrounds often have to make decisions as to the potential correlations between populations' and individuals' hearing characteristics for either an assumed or known on-the-job noise exposure and either an assumed or known off-the-job noise exposure. The recently published ANSI S3.44 standard provides an additional tool in making these types of decisions. This paper will present several case history examples of using ANSI S3.44 in judging the level of protection that had been provided by the company's hearing conservation program for the noise exposed population. The paper will also present the expected changes over time in the hearing of a population or an individual (using a more-likely-than-not judgment criterion).

9:25

3aNS6. Using ANSI S3.44-1996 to assess the hearing losses of individuals. Julia Doswell Royster (Environmental Noise Consultants, Inc., P.O. Box 30698, Raleigh, NC 27622-0698, effective_hcps@compuserve.com)

ANSI S3.44-1996 is a valuable tool for the audiologist to use when reviewing individuals' audiometric records within occupational hearing conservation programs, as well as when evaluating claimants for Worker's Compensation or hearing-related legal actions. By comparing the amount of hearing loss for an individual to the predictions of the model in the standard, the professional can more objectively assess the probable contribution of workplace noise to the individual's total hearing impairment. Likewise, by comparing the time course of the development of loss as shown in hearing conservation program audiograms to model predictions, one can better judge the probability that occupational noise may have caused a recent hearing change (such as an OSHA STS). Examples of actual cases will be shown to illustrate the steps involved in these applications of the standard.

9:35

3aNS7. A proposed method for allocation of compensation for noise-induced hearing loss. Daniel L. Johnson (EG&G MSI, 2450 Alamo SE, Albuquerque, NM 87119-9100)

The method proposed uses the area under a statistical distribution of hearing levels in which the hearing threshold level is above 25 dB. These areas, normalized to 100 individuals, can be calculated for both non-noise-exposed populations as well as populations with various noise exposures. The statistical distributions are calculated by the procedures outlined in ANSI Standard S3.44. One minus the ratio of the non-noise-exposed area to the noise-exposed area provides a fair means for the allocation of hearing loss between noise and presbycusis. This approach considers an individual as representative of a group of exposed workers. This idea was first proposed in Air Force Medical Research Laboratory Technical Report No. TR-80-032, September 1983, Wright–Patterson Air Force Base, OH.

9:45–10:45

Roundtable

WEDNESDAY MORNING, 3 DECEMBER 1997 CHAMBER ROOM, 8:00 TO 9:15 A.M.

Session 3aSAa

Structural Acoustics and Vibration: Active Control of Sound and Vibration

Scott D. Sommerfeldt, Chair
Department of Physics, Brigham Young University, 277 Fletcher Building, Provo, Utah 84602

Contributed Papers

8:00

3aSAa1. Active control of structural intensity in a T-beam. Sabih I. Hayek and Patrick J. Mayfield (Active Vib. Control Lab., Dept. of Eng. Sci. and Mech., Penn State Univ., University Park, PA 16802)

The total vibration structural intensity in a T-shaped connected beam is modeled for a point mechanical noise source located at one end of the horizontal branch of the T-beam. For a point force lying in the plane of the T-beam, only flexural and extensional intensities are included. In order to minimize the total structural intensity propagating into the vertical leg of the T-beam, three control strategies were investigated. These include a point force, a point couple, and a combined point force and couple, where the location of the control actuators were the same for all three strategies. The effectiveness of each strategy is evaluated for control on or off resonances. The total injected power required for control for each control strategy was also evaluated to rate the efficiency of each method.

8:15

3aSAa2. Incorporation of subjective measures in active noise control. Timothy O. Samuels and Scott D. Sommerfeldt (Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602)

An attempt is made to refine a standard active noise-control algorithm in order to take into account the unique response of the human auditory system. It has been shown that decreasing the sound-pressure level at a

location does not guarantee a similar decrease in the perceived loudness at that location [W. R. Saunders and M. A. Vaudry, Proc. of Noise-Con 96, 385–390 (1996)]. Typically, noise cancellation is based on the ''error signal'' from a mechanical device such as a microphone, whose response is nominally flat in the frequency range of the human ear. However, if the response of the ear can be approximated by digitally filtering the error signal before it reaches the adaptive controller, one can, in effect, minimize the more subjective loudness level, as opposed to the sound-pressure level. The work reported here entails simulating noise control based uponminimizing perceived loudness for a varied collection of input noise signals. A comparison of the loudness of the resulting error signal is made to the loudness of the error signal resulting from standard sound-pressure-level minimization. This comparison helps assess the worthiness, for a given noise signal, of utilizing the extra processing time this technique requires.

8:30

3aSAa3. The active control of payload fairing interior acoustics using active blankets. B. H. Houston, M. H. Marcus, and J. A. Bucaro (Naval Res. Lab., Code 7130, 4555 Overlook Ave., S.W., Washington, DC 20375-5350, houston@lpsa1.nrl.navy.mil)

The results of numerical experiments associated with examining the potential performance of active blankets attached to the interior of a payload fairing to reduce the interior acoustic levels are presented. These studies are carried out using a finite-element–infinite-element model of the full fluid-structure system. In this model, the main rocket body is modeled as a finite cylinder with an attached rib-stiffened fairing. The interior fluid is modeled with finite acoustic elements while the entire exterior fluid is modeled with finite and infinite acoustic elements. Examined first is the broadband nature of the interior acoustic response due to a high-order acoustic field developed as a result of a point acoustic source located at the base of the rocket and laterally separated by approximately one rocket length. Next, control simulations are performed that impose a specific acoustic impedance at the interface between the structure and the fluid, thus simulating the acoustic boundary control (ABC) approach. Finally, the performance of an acoustic boundary control implementation is examined for several specific acoustic impedances together with some simple antisound configurations. [Work supported by the Office of Naval Research.]

8:45

3aSAa4. Active structural acoustic control of a thin plate with a hybrid sandwich actuating mechanism. Joshua T. Lee and William Meecham (Mech. and Aerosp. Eng. Dept., 48-121, ENGR-IV, UCLA, Los Angeles, CA 90095)

The objective of this research is to control the acoustic radiation of thin plates with hybrid active–passive actuating mechanisms. The hybrid mechanism consists of a piezoelectric actuator and viscoelastic and piezoelastic active constrained layers in a sandwich configuration, with the viscoelastic core layer in between the piezoelastic components. The actuator end of the hybrid mechanism is attached to a host structure. The main purpose of the proposed hybrid mechanism is to enhance the damping characteristics of the system while maintaining sufficient control authority. The damping characteristic of the hybrid mechanism is maximized by applying opposite voltages to the actuator and the active constrained layers, thus inducing a significant increase in shear strain within the viscoelastic layer. The general goal of this study is to evaluate the performance of the hybrid mechanism with some of the previously published actuating systems. Initial comparison results indicate that the proposed hybrid active–passive sandwich mechanism is more efficient than the actuating devices in comparison.

9:00

3aSAa5. A comparison of an analytical model and experimental results for a curved piezostructure. James K. Henry and Robert L. Clark (Dept. of Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708-0302)

Active structural acoustic control (ASAC) research has previously concentrated on flat plates with attached piezoelectric sensoriactuators. Application of this ASAC research to aircraft interior noise requires the consideration of the curvature of the fuselage panels in the dynamic modeling. This paper investigates the effects of curvature on the dynamics of fuselage panels. A state variable model of a simply supported curved panel with attached piezoelectric sensoriactuators is developed for ASAC applications. The model is verified by comparison to previous work on curved panels and flat plates. Further, the model is validated by experimental data obtained from a simply supported curved panel test structure. [Research supported by NASA Langley Research Center under Grant No. NCC 1 250.]

WEDNESDAY MORNING, 3 DECEMBER 1997 CHAMBER ROOM, 9:30 A.M. TO 12:30 P.M.

Session 3aSAb

Structural Acoustics and Vibration: Vibration of Coupled and Nonhomogeneous Structures

Paul E. Barbone, Chair
Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02115

Contributed Papers

9:30

3aSAb1. Use of genetic algorithms for the vibroacoustic optimization of plates. Alain Ratle and Alain Berry (Département de génie mécanique, Univ. de Sherbrooke, Sherbrooke, QC J1H 1G6, Canada)

Genetic algorithms have proven to be helpful for solving complex optimization problems where more traditional gradient-based methods have failed. The optimal design of vibroacoustic systems often leads to complex optimization problems requiring a method able to find a globally optimal solution through a large number of locally optimal solutions. The vibroacoustic behavior of a simply supported plate carrying one or more point masses has been investigated through a genetic algorithm optimization method, in order to determine optimal placement of the masses for maximal sound or vibration attenuation. It has been shown that the use of an optimization method able to search through the whole optimization domain can give unexpected solutions resulting in a higher level of reduction in radiated noise or vibration response, when compared to intuitive

solutions like placing the point masses on excitation points. In particular, when a well-defined frequency range is targeted, genetic algorithms can be used for tailoring a special design giving highest reduction in the specified range. [Work supported by the IRSST.]

9:45

3aSAb2. End correction in vibrational power flow measurement through joints. Qunli Wu (School of Mech. and Production Eng., Nanyang Technolog. Univ., Nanyang Ave., Singapore 639798, Republic of Singapore)

In the determination of coupling loss factors for complex joints, such as adhesive or bearing joints, analytical methods may not provide satisfactory results. The experimental approaches are normally preferred. In the measurements of the coupling loss factors for the joints, it is assumed that the end reflection can be neglected. However, it is difficult to create such nonreflection termination in the experiments. In this paper, the effects of end reflection in the determination of the coupling loss factors were studied. A new method using the measured reflection coefficients of the termination to correct the measured coupling loss factors was proposed. The experimental study was carried out in the stepped beams and adhesively bonded beams. The experimentally corrected results were compared with the theory.

10:00

3aSAb3. Is the dissipation induced by ''fuzzy'' substructures real or apparent? M. Strasberg (Code 702, David Taylor Model Basin, NSWC, 9500 MacArthur Blvd., West Bethesda, MD 20817-5700)

There continues to be confusion whether the dissipation which can be induced in a master structure by a multitude of small attached substructures is ''real'' or only apparent. It is shown that the dissipation is *real* during steady-state vibration and can cause the master structure to have a larger loss factor than that of the substructures inducing the loss. The dissipation involves the continuous transfer of vibratory power from the master structure to the substructures and its continuous conversion into heat. During transient excitation, however, some of the initial dissipation in the master structure may be only ''apparent,'' involving the storing of some of the transferred power as vibratory energy in the substructures and its cyclical transfer back and forth between master and substructures until eventually dissipated, at a rate which in this case depends on the loss factors of the substructures. Examples of the two cases will be shown.

10:15

3aSAb4. Modal densities of simple dynamic systems. G. Maidanik, K. J. Becker (DTMB, NSWC/CD, 9500 MacArthur Blvd., West Bethesda, MD 20187), and J. Dickey (Ctr. for Nondestruct. Eval., Johns Hopkins Univ., Baltimore, MD 21218)

Modal densities are parameters that participate in the specification of a statistical energy analysis (SEA) model of a complex composed of a number of coupled dynamic systems. The modal density $n(\omega)$ of a dynamic system in this context, is defined by the equation: $(\Delta\omega)n(\omega)=N(\omega)$, where a frequency band $\{\omega,\Delta\omega\}$ is identified by the bandwidth $(\Delta\omega)$ centered at the frequency (ω) and $N(\omega)$ is the number of resonant modes, in this dynamic system, for which the resonance frequencies lie within this frequency band. Usually, it is assumed that the modal density $n(\omega)$ is a continuous and/or a monotonic function of the frequency (ω). Moreover, it is often argued that the modal densities are identical for dynamic systems that possess the same propagation speeds and the same spatial extents, i.e., the same volumes, areas, or lengths. The implication that the assumption and the argument are particularly valid for the higher frequency range, has designated SEA suitable for ''a high-frequency analysis only.'' The modal densities are examined, for a few simple dynamic systems, in an attempt to reassess the assumption, the argument, and the designation.

10:30

3aSAb5. Waves in an acoustic medium with microstructure. Paul E. Barbone and Rebecca B. Shuman (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington Ave., Boston, MA 02215)

Acoustic wave propagation in a medium with microstructure is considered. By microstructure, it is meant that the medium properties change on a length scale that is short compared to an acoustic wavelength. An effective wave equation is derived for the ''long'' scale acoustic field using the method of multiple scales (i.e., homogenization). It is shown that the standard approximation breaks down in a few wavelengths. The effective equation is renormalized in an attempt to obtain a theory, which is valid for long propagation distances. The resulting equation is related to that derived by McDevitt *et al.* (McDevitt *et al.*, ''Dispersive Elastodynamic Models of Periodic Structures Using a High-Order Homogenization Approach,'' Proceedings of Fourth U.S. National Congress on Computational Mechanics, San Francisco, August 6–8, 1997). Asymptotic validity of the equation for finite propagation distances is demonstrated. Over larger distances, solutions of this effective equation exhibits dispersion and attenuation. Strong scattering terms are shown to dominate these effects, however, when the medium is not statistically homogeneous. This asymptotic theory is validated by comparison to reference numerical calculations. [Work supported by NSF.]

10:45

3aSAb6. Acoustic inspection of inertial confinement fusion targets. Thomas J. Asaki (Los Alamos Natl. Lab., MS K764, Los Alamos, NM 87545)

Acoustic cavity resonances were used to examine the interior surface geometry and gas fill pressures of various prototype inertial confinement fusion targets. A representative target consists of a millimeter-sized spherical beryllium shell filled to high pressure with deuterium/tritium (DT) gas at room temperature. Below the triple point of DT the gas forms a solid and redistributes itself symmetrically within the shell through a process known as beta-layering. The thickness of the solid DT layer, and the symmetry and smoothness of each surface must adhere to strict specifications for ignition to occur. Sound-speed measurements at room temperature provide the gas fill pressure and thus the solid DT layer thickness upon cooldown. These measurements rely upon an accurate equation of state. Degenerate-mode splitting of the cavity resonances provides interior surface geometry information. This technique was applied to a variety of deuterium and helium filled shells at pressures up to 356 atm. Several of these shells were manufactured with known interior surface perturbations. Measured mode splitting is compared with theory and the utility of the technique for cryogenic targets is examined.

11:00

3aSAb7. Effect of rib bending on Bloch wave structure of a rib-stiffened plate. Khushi L. Chandiramani (BBN Systems and Technologies, 70 Fawcett St., Cambridge, MA 02138, chandi@bbn.com)

It is well known that flexural vibration on a flat plate stiffened by periodically placed line supports shows stop/passband behavior in frequency, associated with Bloch waves. For a one-dimensional situation, with plate vibration incident normally on the ribs, a simple model in which the ribs exert forces normal to the plate by virtue of their mass often suffices to explain stop bands that occur when the rib spacing d accommodates an integer number of the half-wavelengths of plate vibration. This paper presents results for cases where the ribs present normal forces as well as moments to the plate. Particular attention is paid to the case where the rotary impedance offered by the rib is limited by the rib's own flexibility in bending. At special antiresonance frequencies associated with the rib's own bending motion, the rib rotary impedance is seen to rise sharply and to give rise to stop bands additional to those generated by the rib normal forces.

11:15

3aSAb8. The effect of wave energy trapping in inhomogeneous elastic structure. Victor T. Grinchenko (Inst. of Hydromechanics, 8/4 Zhelyabova St., 252057 Kiev, Ukraine, vgr@ihm.kiev.ua)

In contrast to an acoustical waveguide the distribution of characteristics of normal modes with respect to the cross section in elastic waveguides is strongly dependent on frequency. Such dependence gives rise to several interesting wave phenomena, which take place in inhomogeneous structures. One of them, namely, energy trapping, is quantitatively and qualitatively considered in this report. In terms of the theory of elastic waveguides, this phenomenon manifests itself in an anomaly of the frequency functions of the reflection (transmission) coefficients of energy in the composite waveguides. These functions are calculated for both in-plane and flexural motions in waveguides consisting of two semi-infinite layers with equal thickness. The results are systematized with respect to the ratio of wave impedances of the layers. The frequency bands where the energy trapping is effective are determined. The dependence of these bands on the mechanical characteristics of materials is different for the in-plane and flexural waves. The mechanism of redistribution of energy among the incident and reflected waves is shown. The function of the normal mode with opposite signs of phase and group velocities is illustrated.

11:30

3aSAb9. Detection and classification of the acoustic emission signals by the vector receiver in an elastic environment. Dat H. Tran, Alexsandr J. Krasilnikov, and Igor. L Oboznenko (Acoust. and Acoustoelectron., Dept. of Electron., Ukrainian Natl. Tech. Univ., Kiev, Ukraine 252057)

The novel method of detection and classification of the developing microfissures in elastic environments is proposed. A probability model of the microfissures development is assigned that depends on time and space. The problem is reduced to the solution of the equation of elasticity with the limit and primary condition assigned by it. The rule of the development of the fissures' radius as discrete in time and space or as a random process is assigned. Two variants of the given problem are analyzed: Voltaire's equation or the ordinary differential equation with variable coefficients. The numerical results are given. A six-component vector-phase receiver settled on the unrestricted boundary is used as a receiver. The projection of the vectors of tension and shift are measured and calculated: their mutual correlation is possible. As a result, information about the intensity vector is provided that determines the coordinates and the intensity of the acoustic emission.

11:45

3aSAb10. Aircraft pipeline quality control in a commercial medium. Tannia Yu. Shkvarnytska (Kiev Inter. Univ. Avia, Kiev, Ukraine 252058)

The method of controlling the presence of microcracks in metal pipeline materials for aircraft systems in a commercial medium (on the stage of quality selection of the systems for the specific aircraft model) is proposed. The structure of the echosignal from the cylindrical pipeline in water is investigated in comparison with the pipeline standard section without microcracks. In the presence of microcracks in shell material, changes in the effective Young's modulus and Poisson's ratio due to echosignal structure and the meaning of statistical longitudinal and shear material modules are defined. That effective modulus contains information about microcracks in the pipeline material measuring 1×3 mm (the cylindrical crack). Approximately 50 cm of the length of the main part is allowed to register changes in material. In particular, it is noticed in the frequency range $f(3-10)$ MHz, where the resonance microcrack's peculiarities become apparent. The analytical formulas of effective elasticity modules that contain information about the presence of the many microcracks (including the sole) are obtained. The proposed method can be applied in the case of aircraft pipeline quality control.

12:00

3aSAb11. Windowed displays of broadband impulse response measurements for finite cylindrical shells. Scot F. Morse, Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814), and Gregory Kaduchak (Univ. of Texas, Austin, TX 78713-8029)

Previously interpreted angle-frequency domain displays of impulse response measurements for thick nd thin finite cylindrical shells [S. F. Morse *et al.*, J. Acoust. Soc. Am. **101**, 3043(A) (1997)] reveal backscattering enhancements due to elastic effects on the cylinder over a range of target tilt angles. Presently, these signatures are examined in the angle-time and time-frequency domains. For selected angles, time-frequency analyses are presented as well as synthetic aperture images for selected apertures. From these it is possible to identify the spatial and angular locations of many of the important elastic contributions. Features associated with a meridional ray, as well as helical rays, have the appearance of originating from the back end of the cylinder. One major elastic reponse observed at large tilt angles is found below the coincidence frequency. [Work sponsored by the Office of Naval Research.]

12:15

3aSAb12. Magnetic excitation and acoustic detection of torsional modes of spherical shells in water. Brian T. Hefner and Phillip L. Marston (Phys. Dept., Washington State Univ., Pullman, WA 99164-2814, bhefner@mail.wsu.edu)

The modal structure of an empty nearly spherical metallic shell subject to fluid loading was examined using a novel form of EMAT (electromagnetic acoustic transducer) used previously for shells in air [B. T. Hefner and P. L. Marston, J. Acoust. Soc. Am. **99**, 2594 (1996)]. It was found to be possible to magnetically excite and acoustically detect the high-Q torsional modes of the shell in water using this method. The quality factor, Q, was found to be approximately 1500. The presence of a bias field in addition to the oscillating field of the coil is needed to excite these modes. The bias field was introduced either by the magnetization of the 440C stainless-steel shell itself or by introducing an external field in the case of an unmagnetized shell. It was found that the lowest torsional mode could be excited by an oscillating field at 60 kHz (the mode frequency) and at 30 kHz (half the excited mechanical frequency). The modes were detected by a hydrophone placed near the shell, however, the mechanism for producing this acoustic signal is poorly understood because it depends on some lack of perfect symmetry for the shell. [Work supported by the Office of Naval Research.]

Session 3aSCa

Speech Communication: Individual Characteristics in Speech and Singing

Keith A. Johnson, Chair
Department of Linguistics, Ohio State University, 1712 Neil Avenue, Columbus, Ohio 43210-1298

Contributed Papers

8:00

3aSCa1. Reliable cues to gender and talker identity are present in a short vowel segment recorded in running speech. Michael J. Owren (Dept. of Psych., 224 Uris Hall, Cornell Univ., Ithaca, NY 14853, mjo9@cornell.edu) and Jo-Anne Bachorowski (Vanderbilt Univ., Nashville, TN 37240)

Based on the source-filter model, individual variation in vocal-tract anatomy has long been proposed to influence the acoustic realization of phonetic gestures. It is therefore noteworthy that humans are sexually dimorphic, with individuals being uniquely marked through variation in many anatomical features. Both observations apply to vocal-production structures, and sounds whose characteristics produce a high-fidelity ''off-print'' of underlying anatomy should therefore provide clear cues to gender and identity. Vowels are arguably optimal in this regard, as they reflect details of both source and filter components. This conception was tested using 15 instances of the utterance ''Test n Test'' recorded from each of 50 males and 75 females. Measurements of the vowel in the first ''Test'' were entered in discriminant-function classification. Using only fundamental frequency and a formant-based estimate of vocal-tract length, 1852 of 1875 (99%) tokens were correctly classified by talker gender. Using all variables, 593 of 750 (79%) tokens recorded from males were sorted by talker, while the comparable figures were 729 of 1125 (65%) for females. These results underscore the importance of differences in vocal-tract anatomy for individual variation in speech production and suggest a prominent role of vocalic sounds in indexical cuing. [Work supported by NIMH.]

8:15

3aSCa2. Becoming familiar with a talker from natural and sinusoidal sentences. Sonya Sheffert, David Pisoni (Dept. of Psych., Indiana Univ., Bloomington, IN 47405), Jennifer Fellowes, and Robert Remez (Barnard College, New York, NY 10027)

This report describes the results of a perceptual training study that was designed to explore how listeners learn to categorize novel voices and how knowledge of a familiar voice generalizes to novel utterances. The speech samples from which the listeners learned to identify individuals were of two kinds: Naturally produced English sentences and sinewave replicas of these sentences. The sinewave items were nonspeech tonal patterns that preserved coarse-grained properties of the talker's vocal tract transfer function while eliminating traditional cues to voice quality. Listeners were trained over several days to identify ten talkers from sinewave or natural speech sentences. Knowledge about the talker's voice was then assessed using two generalization tests in which listeners heard a novel set of sentences and identifed the speaker. One generalization test presented sinewave sentences whereas the other presented naturally produced sentences. The results showed that perceptual learning of a talker's voice can occur even when specific acoustic products of vocal articulation are eliminated from the signal, and that this knowledge generalizes to novel utterances produced by these same talkers. [Work supported by NIH-NIDCD.]

8:30

3aSCa3. Effect of subglottal pressure variation in professional baritone singers' voice source. Maria Andersson, Clara Hultqvist (Dept. of Logopedics and Phoniatrics, Huddinge Hospital, The Karolinska Inst., Stockholm and Speech Music Hearing, KTH, Stockholm), and Johan Sundberg (Speech Music Hearing, KTH, S-10044 Stockholm, pjohan@speech.kth.se)

Subglottal pressure (P_s) is the main variable for control of vocal loudness. Measuring its effects on the glottal voice source is complicated, since independent variation of voice control variables is typically difficult for untrained subjects. Professional singers, however, can be assumed to have learned such independent variation. The voice source was analyzed by inverse filtering of the flow signal in five professional operatic baritone singers who, at three pitches, sang a string of [pV:] syllables while performing a diminuendo. Subglottal pressure was measured from the oral pressure during the [p] occlusion. From the flow glottograms the following parameters were determined: Q_{cl} (closed quotient), dU/dt (negative peak amplitude of the differentiated flow glottogram), glottal leakage, ac flow (peak-to-peak flow amplitude), $H1-H2$ (level difference between the two lowest spectrum partials), and glottal compliance (ratio pulse airflow to subglottal pressure). In all subjects, the effect of P_s was similar and systematic for all parameters except glottal leakage. With decreasing P_s, ac flow and dU/dt decreased continuously while compliance increased and $H1-H2$ and Q_{cl}, showing a strong negative correlation, saturated at about 5 dB and 0.5, respectively. The results will be compared to findings from earlier investigations of untrained voices.

8:45

3aSCa4. The recognition of isolated words and words in sentences: Individual variability in the use of sentence context. Ken W. Grant and Philip F. Seitz (Walter Reed Army Med. Ctr., Army Audiol. and Speech Ctr., Washington, DC 20307-5001)

Auditory–Visual (AV) speech recognition is influenced by at least three primary factors: (1) the ability to extract auditory (A) and visual (V) cues; (2) the ability to integrate these cues into a single linguistic object; and (3) the ability to use semantic and syntactic constraints available within the context of a sentence. In this study, the ability of hearing-impaired individuals to recognize bandpass filtered words presented in isolation and in meaningful sentences was evaluated. Sentence materials were constructed by concatenating digitized productions of isolated words to ensure physical equivalence among the test items in the two conditions. Formulae for calculating k factors [Boothroyd and Nittrouer, J. Acoust. Soc. Am. **84**, 101–114 (1988)], which relate scores for words and sentences, were applied to individual subject data obtained at three levels of isolated word-recognition performance approximating 30%, 50%, and 70% correct. In addition, A, V, and AV sentence recognition in noise was evaluated using natural productions of fluent speech. Two main issues are addressed: (1) the effects of intelligibility on estimates of k within individual subjects; and (2) the relations between individual estimates of k and sentence recognition in noise as a function of presentation modality. [Work supported by NIH Grant DC00792.]

9:00

3aSCa5. Individual differences in perceptual closure at the phoneme level. Arthur Boothroyd (CUNY, 33 W. 42 St., New York, NY 10036)

The amount of acoustic information available from the speech signal was reduced by removing all regions of the waveform in which the rms amplitude fell below a threshold value. This removal was accomplished by envelope center clipping followed by attenuation and envelope expansion. The processing was carried out independently in three-frequency bands, which were then recombined. The result was intended to simulate the effects of hearing loss with recruitment. Phoneme recognition in consonant–vowel–consonant words was measured in eight normally hearing subjects as a function of threshold. Subjects, essentially, heard the top x dB of the speech signal, where x varied from 3–30 dB. Significant individual differences of performance were found, covering a range of ± 3.6 dB when measured in terms of the threshold required for a score of 50%. Because all subjects in this experiment had access to the same sensory evidence, the findings support a perceptual closure explanation of individual differences. It may be, therefore, that some of the individual differences of phoneme-level speech perception by the hearing impaired, or by normals listening in noise, also reflect differences in the ability to complete perceptual closure. [Research supported by NIDCD.]

9:15

3aSCa6. Effects of the size and form of the orofacial structure on vowel production. Kiyoshi Honda (ATR Human Information Processing Res. Labs., 2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-02 Japan and Waisman Ctr., Univ. of Wisconsin, Madison, WI 53705), Michiko Hashi., Chao-Min Wu, and John R. Westbury (Univ. of Wisconsin, Madison, WI 53705),

Individual difference of the orofacial structure and its articulatory and acoustic influences have been studied based on the University of Wisconsin x-ray microbeam database [Honda and Wu, J. Acoust. Soc. Am. **100**, 2598(A) (1996)]. This work describes a subsequent statistical analysis on the same datasets. The outlines of the orofacial structure of 20 subjects (10 Japanese and 10 English speakers) were extracted from lateral x-ray scan images and mapped onto a midsagittal reference space defined by the palatal plane and anterior nasal spine. Vowel records in the microbeam data were analyzed to reveal the effect of orofacial morphological difference on vowel articulation and acoustic parameters. In the main results, the size parameters were commonly correlated with the three formant frequencies. In particular, the lower facial height or the size of the mandibular symphysis showed a clear tendency to inversely correlate with the first formant. Among the parameters of form, oral cavity aspect ratio and mandibular symphysis shape thought to be possible factors affecting vowel gestures, which were reflected in the spread of the second formant in all the vowels analyzed. A multiple regression model of speaker characteristics will be discussed.

WEDNESDAY MORNING, 3 DECEMBER 1997

SAN DIEGO/GOLDEN WEST ROOMS, 9:00 A.M. TO 12:00 NOON

Session 3aSCb

Speech Communication: Speech Perception, Speech Aids, Cross Lingusitic and Second Language Studies (Poster Session)

James E. Flege, Chair
Department of Rehabilitation Sciences, University of Alabama, VH 503, Birmingham, Alabama 35294-0019

All posters will be on display from 9:00 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 9:00 a.m. to 10:30 a.m. and contributors of even-numbered papers will be at their posters from 10:30 a.m. to 12:00 noon. To allow for extended viewing time, posters will remain on display until 3:00 p.m.

Contributed Papers

3aSCb1. Contrastive and noncontrastive nonmodal phonation. Barbara Blankenship (Dept. of Linguist., UCLA, Campbell Hall, Los Angeles, CA 90025, blankens@humnet.ucla.edu)

This study explores three manners of vocal fold vibration: modal (the standard mode of vibration), breathy (where the folds are held apart so that the glottis fully closes for only a very small portion of the vibration cycle), and laryngealized (where the folds are held stiffly and vibration is inhibited). Data from languages with contrasting modal and nonmodal vowels were compared with matching data from languages where nonmodal phonation on vowels is not contrastive, but rather is an effect of the preceding consonant. Three acoustic measures were used to determine the nature and duration of nonmodal phonation in Mazatec, Tagalog, Chong, Mpi, and Navajo. In the contrastive cases, nonmodal phonation never persisted through the entire vowel. It was proportionally longer, however, than in the noncontrastive cases. Results offer insights into articulator limitations, perceptually driven contrasts, and gender differences.

3aSCb2. Development of aids to speechreading based on automatic production of cued speech. Louis D. Braida, Paul Duchnowski, David S. Lum, Matthew G. Sexton, and Maroula S. Bratakos (Res. Lab. of Electron., MIT, Cambridge, MA 02139)

In manual cued speech (MCS) a speaker produces hand gestures to resolve ambiguities among speech elements that are often confused by speechreaders. The shape of the hand distinguishes among consonants; the position of the hand relative to the face among vowels. Experienced receivers of MCS achieve nearly perfect reception of everyday connected

speech. MCS has been taught to very young deaf children and greatly facilitates language learning, communication, and general education. A system that can produce cued speech automatically in real time is currently being developed at MIT. Cues are derived by a speaker-dependent HMM speech recognizer that uses context-dependent phone models and presented visually by superimposing animated handshapes on the face of the talker. The benefits provided by these cues strongly depends on articulation of hand movements and on precise synchronization of the actions of the hands and the face. Cue receivers experienced in the reception of MCS can recognize roughly two-thirds of the keywords in cued low-context sentences correctly, compared to roughly one-third by speechreading alone. Ongoing development aims at increasing the accuracy and robustness of the cue recognizer, refining the display, and simplifying the hardware required for use in classroom settings. [Work supported by NIH.]

3aSCb3. Voice offset characteristics for voiceless fricatives and stops. Ailbhe Ni Chasaide and Christer Gobl (Ctr. for Lang. and Commun. Studies, Trinity College, Dublin 2, Ireland)

This paper is concerned with whether the nature and timing of the glottal abduction gesture might differ for voiceless fricatives and stops. The preconsonantal vowel in VCV sequences were examined, where C =/p(:) b(:) f(:) v(:)/ in French (single consonants), Italian, and Swedish (geminates). The speech pressure waveform was inverse filtered using interactive software and the LF voice source model [Fant *et al.*, STL-QPSR **4**/1985, 1–13 (1985)] was matched to the source signal. Results show a long breathy-voiced transition from a preceding vowel to the voiceless fricative, suggesting that glottal abduction occurs early. For the voiceless stops, however, results confirm a previous study [A. Ni Chasaide and C. Gobl, Lang. Speech **36**, 303–330 (1993)] indicating that the timing of glottal abduction is quite variable and language/dialect dependent. A tentative conclusion is that early glottal abduction in voiceless fricatives may be a universal feature, necessitated by the dual production requirements of effecting devoicing with an open vocal tract and achieving the high oral pressure required for generating oral frication. This clearly is not the case for voiceless stops where the precise control appears to depend on language-specific rules. [Work supported by ESPRIT Project No. 6975.]

3aSCb4. Acoustic comparison of the effects of coarticulation on the production of Japanese and American English vowels. James J. Jenkins, Winifred Strange, Kanae Nishi, Brett H. Fitzgerald, Sonja A. Trent, and David H. Thornton (Speech Percept. Lab., BEH 300, Univ. of South Florida, 4202 E. Fowler Ave., Tampa, FL 33620-8200)

Cross-language similarities and differences in the acoustic variability of vowels as a function of speaking ''style'' (citation versus sentences) and phonetic context were explored by comparing the productions of four adult male native speakers each of American English (AE) and Japanese. Multiple instances of the 11 AE vowels /i, ɪ, e͡ɪ, ɛ, æ, ɑ, ʌ, ɔ, o͡ʊ, ʊ, u/ and the 10 Japanese vowels /i, ii, e, ee, a, aa, o, oo, ɯ, ɯɯ/ produced in citation-form bisyllables in /hVba/ and in CVC syllables /bVb, bVp, dVd, gVg, gVk/ imbedded in a carrier sentence were analyzed. Formant values at three temporal locations within the vocalic nucleus (25%, 50%, 75%) of the CVC syllables were compared with ''canonical'' /hVba/ values to determine the amount of ''target undershoot'' and changes in dynamic formant structure as a function of consonantal context. Vocalic durations were measured to determine the extent to which speaking style and consonantal context influenced relative vowel length information in the two languages. Linear discriminant analyses were conducted to quantify the acoustic overlap within and across languages. Implications of these results for cross-language studies of vowel perception will be discussed. [Work supported by NIDCD and ATR.]

3aSCb5. Low bandwidth aids to speechreading based on conveying speech band amplitude envelopes. Sharieff A. Mansour, Philip M. Nadeau, and Louis D. Braida (Res. Lab. of Electron., MIT, Cambridge, MA 02139)

The amplitude envelopes of filtered bands of speech can improve speechreading by normal hearing listeners substantially [Grant *et al.*, J. Acoust. Soc. Am. **89**, 2952–2960 (1991)]. Acoustic signals that convey these envelopes to listeners with severe hearing impairments are currently being evaluated in field and laboratory studies. An initial field study used SiVo aids to derive the amplitude envelope of the octave band of speech centered at 500 Hz and modulate a 200-Hz tone. Of two listeners who used the aids for two months, one benefitted substantially from the amplitude modulated signal and continues to use the aid in preference to conventional hearing aids. In the laboratory study, four normal-hearing listeners compared speechreading supplements derived from one and two filtered bands of speech. Presenting the envelopes of octave bands at 500 and 3300 Hz by modulating the amplitudes of 200- and 240-Hz tones led to keyword scores in IEEE sentences of 74% correct, compared to 76% corrrect when 200- and 500-Hz tones were modulated, and 63% when only a single 200-Hz tone was modulated. Modulating pairs of tones with closely spaced frequencies may thus assist speechreading by listeners with limited hearing. [Work supported by NIH.]

3aSCb6. Intelligibility of bandpass-filtered speech. Ginger S. Stickney and Peter F. Assmann (School of Human Development, Univ. of Texas, Box 830688, Richardson, TX 75083-0688)

Warren *et al.* [Percept. Psychophys. **57**, 175–182 (1995)] reported high intelligibility for sentences processed by one-third octave bandpass filters. The present study examined the contribution of acoustic and linguistic redundancy to the intelligibility of filtered speech. Separate groups of listeners identified sentences or isolated final words from the SPIN (''speech perception in noise'') test. One group heard unfiltered speech, while others heard filtered speech with center frequencies of 750, 1100, 1500, 2100, or 3000 Hz. Gammatone filters with bandwidths of 1 ERB were used to match auditory filter widths. Consistent with Warren *et al.*, the 1500-, 2100-, and 3000-Hz bands were more intelligible than the lower bands. Filtered words in high-predictability (HP) sentences were identified more accurately than those in low-predictability (LP) sentences (mean difference 20%). Isolated filtered words were less intelligible than LP sentences (23% difference). Unfiltered HP and LP sentences and final words were equally intelligible, with scores exceeding 94%. Filtered HP sentences were well identified for the 2100- and 3000-Hz bands, with performance similar to unfiltered sentences. These findings support the hypothesis that context is emphasized when the acoustic signal is degraded, and illustrate the interactive nature of multilevel processing.

3aSCb7. Acoustic characteristics of perceived vowel quantity and quality in English and Norwegian vowels. Anne Nylund and Dawn M. Behne (English Dept., Norwegian Univ. of Sci. and Technol., N-7055 Trondheim, Norway)

In Norwegian and English, as in many languages, vowels can be characterized in terms of distinctive phonological vowel quantity and vowel quality. Norwegian is described as having phonological distinction between long and short vowels, whereas in English vowel quality is generally seen as more distinct. However, both vowel quantity and quality can acoustically be realized in vowel duration and the accompanying characteristic resonance of a vowel. The goal of this project is to examine the perceptual weight of vowel duration and the first two formant frequencies for Norwegians when hearing Norwegian vowels versus English vowels. Six sets of words based on three Norwegian vowel pairs and three English vowel pairs were resynthesized. The words in each set were resynthesized using the members of the pairs as the end points of a matrix with five degrees of vowel duration and five degrees of $F1$ and $F2$ adjustment. Based on judgments by native Norwegian listeners, results to date suggest that native listeners of Norwegian will be more perceptive to differences

in vowel duration than to spectral characteristics. The results for the English and Norwegian vowels will be compared and discussed in terms of native language influence on vowel perception.

3aSCb8. Lateralization effects in distant shadowing. Bertil Lyberg and Åsa Hällgren (Telia Res. AB, Spoken Lang. Processing, S-136 80 Haninge, Sweden)

In the present investigation the lateralization effects for distant shadowers are studied. The subjects listen to spoken utterances presented to them in either the right or the left ear and repeat the utterances simultaneously. The subjects are from previous shadowing experiments known to be distant shadowers, i.e., they repeat stimuli with mean delays of 400 ms or more [Marslen-Wilson, Speech Commun. **4**, 55–73 (1985)]. The time difference between the incoming and the repeated speech is measured at the vowel onsets and the mean value of these differences are used as a measure of shadowing delay. Results from a series of experiments with natural speech with normal fundamental frequency and with constant fundamental frequency will be presented.

3aSCb9. Combining frication and glottal constriction: Two solutions to a dilemma. Ian Maddieson (Dept. of Linguist., Univ. of California, Los Angeles, 405 Hilgard Ave., Los Angeles, CA 90095-1543)

In ejective consonants the airflow generating the sound comes from raising the larynx with the glottis closed. This pistonlike action only moves a small volume of air. Ejective constants are usually stops, and raising the larynx provides sufficient pressure for a salient burst. A handful of languages are described as having ejective fricatives, but how can one sustain a continuant, such as a fricative, using the ejective mechanism? Two languages reported to have sounds of this type are Tlingit (Na-Dene) and Yapese (Austronesian). Acoustic and aerodynamic data on the fricatives in these languages have been collected in recent fieldwork. This demonstrates two different solutions to the problem of combining glottal closure and frication. Tlingit has genuine ejective fricatives, glottal closure overlaps the entire frication duration, and intraoral pressure reaches a higher peak but for a shorter duration than in a pulmonic fricative. The duration of frication appears to be prolonged by narrowing the constriction more than for a pulmonic fricative. In Yapese the frication and glottal constriction are sequenced; intraoral pressure is the same in plain and glottalized fricatives. Both patterns provide clear contrast by solving the problem of combining frication and glottal closure in different ways.

3aSCb10. The locus of the verbal transformation effect: Two tests of the predictions of the Node Structure Theory. Lisa C. Shoaf and Mark A. Pitt (Dept. of Psych., 1885 Neil Ave. Mall, Ohio State Univ., Columbus, OH 43210)

This study explored the verbal transformation effect, a perceptual phenomenon in which participants report hearing illusory words while listening to a recycling word. The Node Structure Theory, or NST [D. G. MacKay, G. Wulf, C. Yin, and L. Abrams, J. Mem. Lang. **32**, 624–646 (1993)], an interactive model of speech perception and production, has been proposed to account for this phenomenon. Two experiments tested NST's claim that verbal transformations have a lexical origin. Experiment 1 examined the influence that lexical information has on the perception of transformations by comparing words and pseudowords. Experiment 2 used the standard (perception-only) task to explore NST's concept of lexical node stability (the idea that lexical items cause fewer VTs than nonlexical items) and its relationship to the number and types of transformations participants perceive. The results suggest that while the locus of the VTE is not lexical, lexical information does influence the types of transformations participants perceive.

3aSCb11. Rime cognate: A new lexical competitor set for spoken word recognition. Shigeaki Amano (NTT Basic Res. Lab., 3-1, Morinosato-Wakamiya, Atsugi, Kanagawa, 243-01, Japan, amano@av-hp.brl.ntt.co.jp)

The rime cognate, which is a set of words containing the same phoneme sequence as the rime of a given word, was proposed as a new lexical competitor set for spoken word recognition. Partial correlation analyses showed that the rime cognate has a much larger contribution to the identification of English spoken words than the neighborhood and the cohort which had been commonly used as lexical competitor sets. The analyses also showed that only parts of the neighborhood and the cohort are reliable as lexical competitor sets in spoken word recognition, and that these parts are included in the rime cognate. This result indicates that the rime cognate integrates the two different competitor sets, the neighborhood, and the cohort. Based on these findings, the rime cognate model of spoken word recognition was proposed.

3aSCb12. A comparison of the acoustic structure of the vowels /i/, /u/, and /a/ in American and Russian mothers' speech to infants and adults. Jean E. Andruski (Eloquent Technol., Inc., 2389 Triphammer Rd., Ithaca, NY 14850, andruski@lightlink.com), Patricia K. Kuhl (Univ. of Washington, Seattle, WA 98195), Inna A. Chistovich, Ludmilla A. Chistovich, Elena V. Kozhevnikova, Victoria L. Ryskina, and Elvira I. Stolyarova (Early Intervention Inst., St. Petersburg, Russia)

Speech input to infants may exert an important influence on the development of language-specific perception by providing infants with information regarding the phonetic and phonological properties of their language. This study compares the acoustic structure of vowels produced by Russian and American mothers conversing with their infants, with the structure of vowels produced by the same women conversing with an adult. Mothers in both countries were asked to play with their infants (aged 2–5 months) using sets of toys whose names contained the target vowels in at least two different consonant contexts (e.g., sheep, bead). They were asked to use the same words in conversation with an adult. All mothers increased the degree of acoustic separation between vowel categories in infant-directed speech by selectively increasing or decreasing $F1$ and $F2$ values. Relative to American mothers [P. K. Kuhl *et al.*, Science **277**, 684–686 (1997)], Russian mothers showed smaller increases in fundamental frequency and less variation in vocal pitch while speaking to their infants. Nevertheless, mothers in both countries produced highly similar changes in acoustic structure, and comparable increases in the size of the acoustic space delimited by the vowels /i/, /a/, and /u/, when speaking to their infants.

3aSCb13. Effects of language experience on speech perception: American and Japanese infants' perception of /ra/ and /la/. Patricia K. Kuhl (Dept. of Speech and Hearing Sci., Univ. of Washington, Seattle, WA 98195), Shigeru Kiritani (Univ. of Tokyo, Tokyo, 113 Japan), Toshisada Deguchi, Akiko Hayashi (Tokyo Gakugei Univ., Tokyo, 184 Japan), Erica B. Stevens, Charmaine D. Dugger, and Paul Iverson (Univ. of Washington, Seattle, WA 98195)

Listening to language during the first year of life has a dramatic effect on infants' perception of speech. With increasing exposure to a particular language, infants begin to ignore phonetic variations that are irrelevant in their native language. To examine these effects, 72 American and Japanese infants were tested at two ages, 6–8 months and 10–12 months, with synthetic versions of the American English /r/ and /l/ consonants. The /r–l/ contrast is not phonemic in Japanese. In both countries, the same experimenters, technique (head-turn conditioning), and stimuli were used. The results revealed two significant effects. The first shows the impact of language experience on speech perception. At 6–8 months of age, American and Japanese infants did not differ. Both groups performed significantly above chance (American $M=63.7\%$; Japanese $M=64.7\%$). By 10–12 months of age, American infants demonstrated significant improvement relative to performance at 6–8 months ($M=73.8\%$), while Japanese in-

fants declined ($M = 59.9\%$). Second, performance varied significantly as a function of the direction of stimulus change (/l/ to /r/ easier than the reverse), regardless of age or language experience. Discussion will focus on separating effects attributable to linguistic and psychoacoustic factors. [Work supported by NIH.]

3aSCb14. Exploitation of stress information in spoken-word recognition in Dutch. Wilma van Donselaar and Anne Cutler (MPI for Psycholinguist., P.O. Box 310, 6500 AH Nijmegen, The Netherlands)

A fragment-priming experiment addressed the question of how stress information is exploited in the process of recognizing spoken Dutch words. Materials were 20 pairs of Dutch words such as *octopus/oktober*, with two initial syllables which were segmentally identical but different in stress; both begin *okto-* (N.B.: in neither case is their vowel reduction, in Dutch), but *octopus* is stressed on the first syllable, *oktober* on the second. Exploiting the known effect that visual lexical decisions are made more rapidly if subjects have just heard the word, or part of it, we compared the lexical decision response times of 56 native speakers of Dutch to the target words (e.g., *OKTOBER*), presented visually immediately following a spoken presentation of a neutral sentence ending with a partial word. This partial word had (a) a correctly stressed beginning (e.g., *okTO-*), (b) an incorrectly stressed beginning (e.g., *OKto-*), or (c) a control beginning (e.g., *eufo-*). Responses were significantly faster (than the control condition) only after correctly stressed primes; that is, *OKto-* primed *OCtopus* but not *okTOber*, *okTO-* primed *okTOber* but not *OCtopus*. Thus Dutch listeners can exploit stress information effectively at a relatively early stage in the spoken presentation of a word.

3aSCb15. An examination of word–initial-stop closure interval in English, Spanish, and Spanish–English bilinguals. Kerry P. Green (Dept. of Psych., Univ. of Arizona, Tucson, AZ 85721, kgreen@u.arizona.edu), Mary L. Zampini, and Joel Magloire (Univ. of Arizona, Tucson, AZ 85721)

This study investigated the effect of the linguistic experience on the duration of the preceding closure interval (CI) on word–initial-stop consonants. Native speakers of English (NE), Spanish (NS), and Spanish–English bilinguals produced sentences containing words beginning with either a voiced or a voiceless stop consonant. As is typical for word–initial stops in English, no difference in the CI occurred between voiced and voiceless consonants for the NE speakers. The NS speakers produced the voiceless stops with CIs similar to the NE voiced stops (consistent with the fact that both are classified phonetically as short-lag stops). The voiced stops, however, had significantly shorter CIs. Like the NE speakers, the bilinguals in the English mode produced voiced and voiceless stops with equal CIs. In the Spanish mode, the bilinguals maintained a distinction in CI between voiced and voiceless stops, although the CIs were significantly different from their NS counterparts. The results suggest that Spanish speakers use CI to help distinguish the voicing characteristics of stops in word–initial position and that bilinguals reweigh different voicing cues as a function of the language mode.

3aSCb16. The effect of masked syllable primes on word and picture naming. Niels O. Schiller (Max Planck Inst. for Psycholinguist., NL-6500 AH Nijmegen, The Netherlands)

The role of the syllable in Dutch speech production is investigated. Four experiments tested the effect of masked syllable primes on the naming latencies of words and pictures. Target words were bisyllabic Dutch nouns that either had clear syllable boundaries and began with a CV syllable (e.g., fa.kir ''id.'') or a CVC syllable (e.g., fak.tor ''factor''), or their syllable boundary was ambiguous, in which case they began with a CV[C] syllable (e.g., fa[kk]el ''torch''). In the syllable match condition, targets were preceded by syllable primes that were identical to the first syllable. In the syllable mismatch condition, the syllable prime was one segment shorter or longer than the target word's first syllable. A neutral condition was designed to determine the direction of the priming effects (facilitation or inhibition). All related primes facilitated the naming of the targets significantly, but the priming effect was independent of the syllabic structure of prime and target. It is concluded that the syllable does not play a functional role in the process of phonological encoding in Dutch. Since the size of the facilitation effect increased with increasing overlap between prime and target, the priming effect is accounted for by a segmental overlap hypothesis.

3aSCb17. Abstract withdrawn

3aSCb18. A cross-linguistic study of diphthongs in spoken word processing in Japanese and English. Kiyoko Yoneyama (Dept. of Linguist., Ohio State Univ., 1712 Neil Ave., Columbus, OH 43210, yoneyama@ling.ohio-state.edu)

Utilizing a phoneme-monitoring task, the current study investigates the sensitivity to moraic structure in English and Japanese diphthongs by three groups of language users: monolingual Japanese listeners, monolingual English listeners, and semibilingual Japanese speakers of English. Experiment 1 focused on monolingual Japanese listeners and found that they did not show a moraic effect in English materials, although they did not show that in Japanese materials. Experiment 2 focused on monolingual English listeners and found that they were not sensitive to moraic structure in either English or Japanese, and seemed to listen to both English and Japanese in the same listening strategy. Experiment 3 focused semibilingual Japanese speakers of English and found that they showed a moraic effect in Japanese materials while they did not in English materials. Collectively, findings in experiments 1 and 3 suggest that Japanese natives are generally sensitive to moraic structure. Also, those in experiment 1 and 2 suggest that sensitivity to moraic structure is language specific. Finally, those in experiment 3 suggest that extensive second-language experience allows one to readjust this tuning for each language spoken, such that a listener's speech segmentation strategy can differ by language.

3aSCb19. Lexical influences on the McGurk effect. Lawrence Brancazio (Haskins Labs., 270 Crown St., New Haven, CT 06511 and Univ. of Connecticut, Storrs, CT 06269, lab93006@uconnvm.uconn.edu)

Much research has been devoted to the study of whether lexical knowledge affects lower-level phonetic processing. For example, in the Ganong effect (Ganong, JEP:HPP **6**, 110–125), the lexical status of the end points of a /b/–/p/ voicing continuum affects the point at which a shift from /b/ to /p/ responses is made (e.g., more /b/ responses in a /bif/–/pif/ continuum than in a /bis/–/pis/ one). This study addresses whether lexical effects influence the perception of discrepant audiovisual stimuli (the McGurk effect). Pairs of words and nonwords were constructed that differed only in the place of articulation of the initial phoneme; in each pair either both members were words, only one was a word, or neither was a word. The stimuli were presented audiovisually, with the audio from one pair member and the video from the other; subjects identified the intial phoneme. Analyses of a proportion of video responses (McGurk responses) indicated significant effects of the lexical status of the auditorily-presented token and of the visually-presented token: the McGurk effect was stronger when the video token was a word, and weaker when the audio token was a word. Implications of these findings will be discussed.

3aSCb20. Effects of identification training on Japanese adults' perception of American English vowels. Winifred Strange (Speech Percept. Lab., BEH 255, Univ. of South Florida, 4202 E. Fowler Ave., Tampa, FL 33620-8150) and Reiko Akahane-Yamada (ATR Human Information Processing Res. Labs., Kyoto, Japan)

In a pre-test, 20 Japanese listeners' identification of American English vowels /ɛ, æ, ɑ, ʌ, ɔ/ was examined, using syllables /bVb, bVp, dVd, dVt, gVg, gVk/ embedded in sentences produced by four men (CVCs) and real words produced by two men and one woman (Words). Experimental subjects (N=10) completed nine training sessions on /æ, ɑ, ʌ/, using CVCs (three speakers). Pre-test-Post-test comparisons showed that identification of trained vowels in CVCs improved for the experimental group (42% to 85% correct), but not for the control group (44% to 44%). The training effect generalized to the speaker not used in training. Experimental subjects also improved significantly on trained vowels in words (43% to 63%); the control group did not (44% to 47%). Again, the training effect transferred to novel speakers. Accuracy on untrained vowels /ɛ, ɔ/ decreased somewhat for the experimental group, while it improved for the control group, suggesting the presence of response biases. Nevertheless, overall greater performance gains for the experimental group (23% for CVCs, 10% for words) versus the control group (6% for CVCs, 6% for words) suggest that robust changes in perception of non-native vowels are possible with identification training. [Work supported by NIDCD and ATR.]

3aSCb21. Effects of audio-visual training on the identification of English /r/ and /l/ by Japanese speakers. Reiko Akahane-Yamada (ATR Human Information Processing Res. Labs., Hikaridai, Seika, Soraku, Kyoto, 619-02 Japan), Ann R. Bradlow, David B. Pisoni (Indiana Univ., Bloomington, IN 47405), and Yoh'ichi Tohkura (ATR, Kyoto, 619-02 Japan)

Two groups of Japanese speakers were trained to identify AE /r/ and /l/ using two different types of training: audio visual and auditory only. In audiovisual training, a movie of the talker's face was presented together with the auditory stimuli, whereas only auditory stimuli were presented in audio-only training. Improvement in /r/–/l/ identification from pretest to post-test on three types of tests (audio-only, visual only and audiovisual) did not differ substantially across the two training groups. Interestingly, the audio-only group showed improved identification in the visual-only tests, suggesting that training in the auditory domain transferred to the visual domain. A McGurk-type test using /r/ and /l/ stimuli with conflicting audio and visual information was also conducted. Identification accuracies on this test showed a greater effect of conflicting visual information at post-test than at pretest for the audio–visual training group, but not for the audio-only training group, suggesting that audio-visual training facilitated integration of auditory and visual information. Taken together, these results suggest that the internal representation of second-language phonetic categories incorporates both auditory and visual information. Implications for theories of perceptual learning and phonological development will be discussed.

3aSCb22. Communicative factors in speech variation: Hyperarticulatory compensation for lexical competition. Richard Wright (Speech Res. Lab., Dept. of Psych., Indiana Univ., Bloomington, IN 47405)

Based on factors in lexical competition such as usage frequency and similarity-neighborhood density, previous research has characterized words as ''easy'' or ''hard'' to identify. Lindblom [in W. Hardcastle and A. Marchal (Eds.), 403–439 (1990)] has proposed that talkers accommodate listeners' communicative needs by controlling the degree of reduction (hyper- and hypoarticulation) in different contextual conditions. This study examines the degree of centralization of vowels in 38 ''easy'' and 38 ''hard'' monosyllabic (CVC) words of equal familiarity spoken in isolation by ten talkers. Measurements of the first two vowel formants ($F1$, $F2$) are made at the point of maximal displacement in the vowel (excluding the initial and final 50 ms of the vowel). Centralization is measured by calculating the Euclidean distance from the center of a talker's $F1$–$F2$ vowel space. Three results emerge: (1) overall ''easy'' words were significantly more centralized than ''hard'' words; (2) peripheral vowels, such as /i/, /a/, /u/, showed the greatest effect; and (3) there was considerable between-talker variability both in overall vowel centralization and in the magnitude of the difference between vowels in ''easy'' and ''hard'' words. [Work supported by NIH Grant No. DC00012.]

3aSCb23. Computational approaches to relating consonant and sentence recognition test scores. Philip F. Seitz and Ken W. Grant (Army Audiol. & Speech Ctr., Walter Reed Army Med. Ctr., Washington, DC 20307-5001, seitz@wrair-emh1.army.mil)

Several recent studies have investigated the predictive relations among phoneme, word, and sentence recognition tests [e.g., Olsen *et al.*, Ear Hear. **18**, 175–188 (1997)]. The present study addressed this issue computationally, using data from 29 older hearing-impaired subjects who performed audio–visual recognition of 18 consonants in VCV syllables, and key words in IEEE/Harvard sentences. It was hypothesized that the predictive relation between consonant and sentence test scores should be made stronger by explicitly modeling and adjusting for differences between the test materials. Two approaches were tested: (1) weighting scores for the consonants based on their frequency of occurrence in the sentence materials, and (2) using a large computerized English lexicon to obtain estimates of possible lexical confusions for the IEEE/Harvard key words based on the consonant confusion matrices. The coefficient of determination (r^2) for the overall consonant and key word scores was 0.56 [Grant *et al.*, J. Acoust. Soc. Am. **97**, 3308 (1995)]. Neither the frequency weighting nor the lexical weighting of the consonant scores increased the r^2. These results suggest that consonant and sentence recognition performance are related by a very general property of speech recognition, which is not subject to lexical mediation. [Work supported by NIH-NIDCD.]

3aSCb24. Accuracy in producing and perceiving vowels in a second language (L2). James Emil Flege (Dept. of Rehabilitation Sci., Univ. of Alabama, VH 503, Birmingham, AL 35298), Ian R. A. MacKay (Univ. of Ottawa, East, Ottawa, ON K1N 6N5, Canada), and Diane Meador (Univ. of Alabama, Birmingham, AL)

The aim of this study was to assess the effect on L2 performance of native Italian subjects' age of arrival (AOA) in Canada and their percentage use (% USE) of Italian. The 72 native Italian subjects examined had lived in Canada for 36 years, on average. They repeated auditorily modeled English words containing 1 of 11 vowels, then inserted each vowel into a nonword (/b__do/) context. Analyses of listeners' goodness ratings revealed that subjects with AOAs greater than 10 years produced vowels less accurately than did native English controls. These same subjects produced four English vowels that are not found in Italian (/ɪ æ ʊ ɚ/) even less accurately in nonwords than in words. This suggested that the ability to establish categories for ''new'' L2 vowels is age-limited. AOA and % USE were also found to influence the native Italian subjects' performance on an oddity-format categorial discrimination test which examined contrasts between English vowels (e.g., /ɪ/–/i/) or English versus Italian vowels (e.g., /æ/–/a/). Production and perception accuracy were found to correlate significantly. Taken together, the results support the view that L2 vowel production accuracy is limited by the ability to establish new categories perceptually which, in turn, is age-limited. [Work supported by NIH.]

3aSCb25. Resolution of lexical ambiguity by emotional tone of voice. Lynne C. Nygaard (Dept. of Psych., Emory Univ., Atlanta, GA 30322)

The present study was designed to evaluate effects of emotional tone of voice on perception of word meaning. In two experiments, listeners were presented with emotional homophones that had one affective mean-

ing (happy or sad) and one neutral meaning and were spoken in either a happy, sad, or neutral tone of voice. In the first experiment, listeners were asked to make tone of voice judgments for homophones that had been low-pass filtered (condition 1) or were presented unaltered (condition 2). The results showed that the affective meaning of the homophone affected tone of voice judgments, but only when homophones were not low-pass filtered. When presented unaltered, tone of voice identification was more accurate when affective tone and meaning were congruent. In the second experiment, listeners were asked to transcribe the emotional homophones in three tone of voice conditions: happy, neutral, and sad. The results showed that listeners provided more affective transcriptions when the tone of voice was congruent. That is, listeners reported more happy transcriptions when the tone of voice was happy and more sad transcriptions when the tone of voice was sad. These findings suggest that emotional tone of voice can affect the selection of word meaning.

3aSCb26. The relation between the age of learning a second language (L2) and the degree of foreign accent. Hua Liu, James E. Flege (Dept. of Rehabilitation Sci., Univ. of Alabama, VH503, Birmingham, AL 35298), and Grace H. Yeni-Komshian (Univ. of Maryland, College Park, MD)

A 1995 study by Flege, Munro, and MacKay [J. Acoust. Soc. Am. **97**, 3125–3134 (1995)] examined the degree of perceived foreign accent in English sentences spoken by Italians whose ages of arrival (AOA) in North America ranged from 2 to 22 years. Two surprising findings were obtained: (1) A linear relation was found to exist between AOA and the degree of perceived foreigen accent; and (2) many of the native Italian subjects who began learning English as young children were found to have detectable foreign accents. To determine if these findings were generalizable, a comparable population of native Korean subjects was examined. English sentences spoken by 240 Koreans and 24 native English (NE) controls were rated for degree of foreign accent by ten native English-speaking listeners. As in the earlier study, a linear relation was found to exist between AOA and degree of accent; and sentences spoken by Koreans who arrived in the United States as young children received significantly lower ratings than those spoken by the NE controls. It appeared, however, that the native Korean subjects generally had stronger foreign accents than did the native Italians subjects. Possible explanations for this cross-language difference will be discussed. [Work supported by NIH.]

3aSCb27. Pronunciation proficiency in L1 and L2 among Korean–English bilinguals: The effect of age of arrival in the U.S. Grace H. Yeni-Komshian (Dept. of Hearing and Speech Sci., Univ. of Maryland, College Park, MD 20742, gykomshi@bssl.umd.edu), James E. Flege, and Hua Liu (Univ. of Alabama, VH503, Birmingham, AL 35294)

Pronunciation proficiency in the first (L1=Korean) and the second (L2=English) languages of bilinguals who had emigrated to the U.S. was examined. The focus of this presentation is on the effect of age of arrival (AOA) in the U.S. on pronunciation in L1 and in L2. The subject pool consisted of 240 Korean–English bilinguals classified into ten subgroups (ages 2–22) based on AOA. Two groups of monolingual subjects were also tested: 24 Korean and 24 English speakers. The bilingual subjects produced sentences in their L1 and L2 and the monolingual subjects produced sentences in their L1's. Native listeners from each language group rated the speakers' pronunciation on a 9-point scale. The results show a positive correlation between AOA and Korean pronunciation (0.74) and a negative correlation between AOA and English pronunciation (-0.85). In each language group and regardless of AOA, the monolingual subjects were rated better in pronunciation than the bilingual subjects. The implications of these findings to second language acquisition and first language retention will be discussed. [Work supported by NIH.]

3aSCb28. The influence of prosodic phrasal boundaries on the resolution of semantic lexical ambiguity. Amy J. Schafer and Shari R. Speer (Speech-Lang.-Hearing Dept., 3031 Dole Ctr., Univ. of Kansas, Lawrence, KS 66045, schafer@ukans.edu)

Two experiments explored the effect of two levels of prosodic phrasing, the intonational phrase (IPh), and phonological phrase (PPh), on the interpretation of ambiguous words in sentences. In experiment 1, polysemous words were presented in semantically neutral sentence-initial clauses, followed by either an IPh or PPh boundary. A second clause resolved the ambiguity. End-of-sentence judgment times showed that reanalysis to the subordinate meaning of the ambiguous word took longer following IPh than PPh boundaries, suggesting that more extensive interpretive processing had taken place following the higher-level boundary. In experiment 2, strongly-biased polysemous words were presented in clauses weakly biased toward their subordinate interpretation, followed by either an IPh or PPh boundary. Following IPh boundaries, visually presented semantic associates of the subordinate meaning of the polysemous word were named faster than targets associated with the dominant meaning. This difference was not found for PPh conditions. Phonological and phonetic analyses will be presented. It will be argued that these processing effects cannot be explained solely on the basis of durational differences between IPh and PPh boundaries. The results support the view that the human sentence processor builds a prosodic representation and is guided by it during semantic interpretation. [Work supported by NIMH.]

3aSCb29. Speech perception in an individual with verbal auditory agnosia. Q. Emily Wang, Richard K. Peach (Dept. of Commun. Disord. and Sci., Rush Univ., 1653 W. Congress Pkwy., Chicago, IL 60612), Yi Xu (Northwestern Univ., Evanston, IL 60208), and Charles Manry II (Rush Univ., Chicago, IL 60612)

Verbal auditory agnosia or pure word deafness is a clinical disorder when an individual is able to comprehend written language but unable to comprehend spoken language in the absence of hearing difficulties. It is due to brain damage involving the primary auditory cortex or connections between the thalamus and this area. Although it is a well-known fact that an individual suffering from unilateral auditory agnosia is unable to comprehend speech, it is not clear whether his/her speech acoustic processing as a whole or different aspects of this processing are affected. In an attempt to examine this issue, speech perception ability was tested in an individual who presented with typical clinical behaviors of unilateral verbal agnosia as a result of brain damage involving cortical and subcortical white matter of the left temporal lobe. A series of speech discrimination and identification tests were conducted at two time points: 1 week post onset and 1 month post onset. The testing results indicated that aspects of speech acoustic processing may be differentially affected in unilateral verbal agnosia. [Work supported by Rush University and Northwestern University.]

3aSCb30. Listeners' representations of within-word structure for bilingual Japanese speakers of English and monolingual speakers of Japanese and English. Takashi Otake (Dept. of English, Dokkyo Univ. and Dept. of Linguist., Ohio State Univ., 220 Oxley Hall, 1712 Neil Ave., Columbus, OH 43210-1298, otake@dokkyo.ac.jp) and Keiko Yamamoto (National Rehabiliation Ctr. for the Disabled, College, Namiki-cho, Tokorozawa, 359, Saitama, Japan)

Assuming that listeners may form conscious representations of potential within-word structure, it was investigated what phonological units could be exploited by monolingual speakers of Japanese and English

[Otake *et al.*, Proc. EUROSPEECH **95**, 1703–1706 (1995)]. The present study investigated how bilingual Japanese speakers of English and monolingual speakers of Japanese and English could form conscious representations of potential within-word structure in the two languages. The three groups of subjects were presented with spoken words in both languages and asked to mark on a written transcript of each word (e.g., buranko for Japanese and veranda for English) the second natural division point from the onset in the word. The statistical analysis showed that both the Japanese and English monolinguals exploited morae and syllables, respectively, and that the bilinguals exploited syllables. These results suggest that the exploitation of phonological units to represent within-word structure may depend upon the degree to which the structure of the language being acquired encourages the use of the units and that the system of the representation may be highly compatible with speech segmentation procedures [Cutler *et al.*, Cogn. Psychol. **24**, 381–410 (1992)]. [Work supported by a Fulbright senior research grant and TAF.]

WEDNESDAY MORNING, 3 DECEMBER 1997 FRIARS/PADRE ROOMS, 8:00 TO 11:30 A.M.

Session 3aSP

Signal Processing in Acoustics, Psychological and Physiological Acoustics and Speech Communication: Acoustics in Multimedia III: Source Localization

Monica L. Hawley, Chair
Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215

Invited Paper

8:00

3aSP1. Three experiments in aurally guided visual search. Richard L. McKinley, William R. D'Angelo, W. Todd Nelson (Armstrong Lab., Wright–Patterson AFB, OH 45433-7901), David R. Perrott (California State Univ., Los Angeles, CA 90032), and Robert S. Bolia (Systems Res. Labs., Dayton, OH 45440-3696)

Aurally guided visual search is a logical concept if one considers the evolutionary utility of the ability to localize sound sources, such as food or predators by primal man. The localized auditory stimulus would direct the gaze to the approximate location of the target or threat. The auditory system has a full 4π sr field while the visual system is limited to about 150 deg horizontally by about 60 deg vertically. The spatial resolution of the visual system is high, while the auditory system's spatial resolution is low. A series of experiments is being conducted investigating the effects of spatial auditory cueing on visual search. This paper describes the design, methodology, and results of the first three aurally guided visual search experiments in this series. The experiments were conducted using both real and virtual spatial auditory stimuli and included visual target location acquisition and discrimination in a black visual field, visual target detection in a simple gray visual field, and visual target location acquisition and discrimination in visual fields of varying complexity. The results indicate that spatial auditory cues significantly improve performance in a visual search task in all visual environments investigated. [Work supported by AFOSR.]

Contributed Papers

8:30

3aSP2. Aurally aided visual search in depth. Raymond Lim, David R. Perrott, and Jose A. Gallegos (Dept. of Psych., California State Univ., 5151 State Univ. Dr., Los Angeles, CA 90032)

Minimum latencies required to locate and identify a visual target (visual search) were obtained in which the visual target could appear in a number of places on a depth plane (0-deg azimuth). A visual target was among the visual distracters located on a finite array segregated at three distances with placement of the target varied within the session from 24 to 232 in. from the subject. The stimulus was a broadband noise (70 dB A). Factoring in the distance of each speaker in relation to the subject, the inverse square law was considered in this paradigm. Four subjects were tested in three conditions: aurally aided search (spatially correlated), aurally unaided search (spatially uncorrelated, where no spatial information was provided), and unaided search (no auditory stimulus was used). The introduction of aurally aided search cues enhanced the ability to localize a target [$f(38,84)=2.11$, $p<0.005$], thus lessening the workload on the visual system when distracters were in the visual field. Theoretical aspects and applications will be discussed further.

8:45

3aSP3. Aurally aided visual search: The effects of individualized head-related transfer functions on response times. Robert S. Bolia (Systems Res. Labs., Dayton, OH 45440, rbolia@biocom.al.wpafb.af.mil) and Richard L. McKinley (Wright–Patterson AFB, OH 45433)

Previous experiments by the authors have investigated the effects of spatially correlated auditory cues on visual search times. In a recent study [D'Angelo *et al.*, J. Acoust. Soc. Am. **101**, 3107 (1997)], it was shown that the addition of a spatially correlated auditory cue, presented over headphones, effects a significant reduction in response times for a search for a single visual target among a varying number of similar distractors. The aim of the present study is to investigate the effects of individualized head-related transfer functions (HRTFs) on response times in an aurally aided visual search task. Subjects will search for a visual target among 1, 5, 10, 25, or 50 distractors with the aid of a spatially correlated auditory cue, using either their own or generic (90th percentile KEMAR) HRTFs. Results will be presented and the implications for the design of virtual auditory systems discussed. [Work supported by AFOSR.]

9:00

3aSP4. Source identification—Theory and experiment. William M. Hartmann, Brad Rakerd, and Joseph B. Gaalaas (Michigan State Univ., East Lansing, MI 48824)

A source identification experiment measures the ability of a listener to identify the location of a sound source. The experimental results are in the form of an rms error and a variability. The experiment can be described in terms of a statistical model in which the listener's responses are determined by the width and bias of an internal distribution. Model calculations were performed to aid the design of source identification experiments. Calculations in the limit of small spans ($\ll$180 deg) showed that the experimental variability provides a good estimate of the width of the internal distribution if both the bias and the source spacing are not greater than the width. The results are insensitive to the number of sources in the array. Specific models were developed for dealing with large biases, a frequent experimental occurrence. Calculations in the limit of large spans showed that experiments that ignore front-to-back reversals may lead to misleading results. Experimental tests of the models, using four human listeners in experiments having 3, 6, 12, or 24 sources, generally corroborated the models for both easy and difficult localization conditions. [Work supported by the NIDCD, DC00181.]

9:15

3aSP5. Three-dimensional auditory localization of nearby sources. Douglas S. Brungart, Nathaniel I. Durlach, and William M. Rabinowitz (Res. Lab. of Electron., MIT, 77 Massachusetts Ave., Cambridge, MA 02139, brungart@mit.edu)

Auditory localization performance has been measured for a point-source stimulus within 1 m of a listener's head. Broadband noise bursts, roved in intensity to eliminate loudness as a distance cue, were presented at random locations ranging from 0°–180° in azimuth, 60°–60° in elevation, and 15 cm–1 m in distance. The listeners responded to these stimuli by moving a response wand to the perceived location of the sound. Response errors were then calculated in azimuth, elevation, and distance. Directional localization error for nearby sources was comparable to that previously reported in localization experiments at distances greater than 1 m. Distance accuracy was found to depend on source azimuth, with best performance to the listener's side and worst performance directly in front of the listener. Distance judgments were slightly more accurate when the amplitude of the stimulus was not randomized to eliminate distance effects. The improvement in accuracy for the fixed-amplitude stimulus was greater in front than to the side of the listener. Distance accuracy was substantially degraded when one of the listener's ears was occluded by an earplug and muff. These results suggest the importance of binaural cues to near-field distance perception. [Work supported by AFOSR.]

9:30

3aSP6. Improving the acuity and robustness of localization of multiple sources using ''stencil'' filters. Chen Liu, Albert S. Feng, William D. O'Brien, Jr., Bruce C. Wheeler, Robert C. Bilger, and Charissa R. Lansing (Beckman Inst., Univ. of Illinois at Urbana–Champaign, Urbana, IL 61801, liu2@uiuc.edu)

A biologically inspired localization and speech-enhancement technique, using two microphones, was recently developed for extracting and characterizing a sound in the presence of multiple interfering sources [Liu *et al.*, The Second Biennial Hearing Aid R&D Conference (1997)]. In the previous study, a ''stencil'' filter was created in the frequency versus intermicrophone phase delay space which significantly enhanced the acuity of sound localization. The stencil followed the multiple curves which represent phase coherence at multiple time differences. However, source artifacts resulted from the curved lines that pose a problem for identifying real versus artifactual sources and for obtaining accurate estimates of sound azimuths. These problems are more severe when the number of sources is large and when some of the sources are less intense than the others. The current study further investigated the role of the curved lines in localization from a *weighted* stencil filter using different weighting methods. Furthermore, an improvement was made in the localization procedure. By adjusting the weights of the stencil filter, the source artifacts were minimized and the acuity and robustness of localization were increased. The improved algorithm was particularly effective in localizing weak sound sources and in situations having a large number of sources.

9:45–10:00 Break

10:00

3aSP7. Human localization of sound signals with reduced bandwidth. Flemming Christensen, Morten Lydolf, and Michael F. Soerensen (Acoust. Lab., Aalborg Univ., Fredrik Bajers Vej 7 B4, DK-9220 Aalborg, Denmark, acoustics@kom.auc.dk)

In multimedia systems, the sound is often presented with reduced bandwith. As 3-D sound is introduced, this might give rise to localization problems, a consequence which is partly confirmed by this investigation. A listening experiment was made, in which 12 subjects participated. The subjects were placed in an anechoic room in a setup with 17 loudspeakers placed in different directions. As a reference experiment the subjects listened to a pink noise signal played from one loudspeaker at a time. The subjects were asked to point out the loudspeaker from which they perceived the sound. In eight other experiments the sound was low-pass filtered at frequencies of 1, 2, 4, or 8 kHz, or high-pass filtered at 0.5, 1, 2, or 4 kHz. The procedure was the same as in the reference experiment. For each subject the order of the tests was randomized. The experiment showed that all four low-pass filtered signals gave an increase in localization errors. This was not the case with the high-pass filtered signals.

10:15

3aSP8. The contribution of interaural intensity differences to the horizontal auditory localization of narrow bands of noise. Matthew H. Bakke (Lexington School for the Deaf/Ctr. for the Deaf, Res. Div., 30th Ave. and 75th St., Jackson Heights, NY 11370) and Harry Levitt (CUNY Graduate School, New York, NY 10036)

Brief noise bursts (340-ms, 20-ms rise/fall times) consisting of third-octave bands of noise (center frequencies 0.5, 1, 2, and 4 kHz) and band-pass noises with different degrees of low-frequency content (0.5–4, 1–4, and 2–4 kHz) were recorded binaurally from 17 different locations (90 deg on the left to 90 deg on the right in 11.25-deg steps) at the ears of KEMAR in an anechoic room and a reverberant room. The recorded sounds were processed by attenuating or removing interaural intensity differences (IIDs) and presented to five normally hearing subjects through insert transducers (ER-3A) in a sound-source identification task. The effect on localization accuracy of reducing IIDs by half was not significant. However, reducing IIDs to zero resulted in degraded localization accuracy for noise bands that contained energy above 1000 Hz. The results were similar for signals recorded in a reverberant room as for those recorded in an anechoic room. Results will be discussed with respect to the relative contributions of interaural time differences (ITDs) and IIDs for horizontal auditory localization. [Work supported in part by NIH, NIDRR.]

10:30

3aSP9. Audible apparent movement in depth. Jose A. Gallegos, David R. Perrott, Raymond Lim, and Agavni Petrosyan (Dept. of Psych., California State Univ., 5151 State University Dr., Los Angeles, CA 90032)

Auditory apparent movement was studied for source distance in depth varying in both relative and absolute distance. The stimuli consisted of 50-ms broadband noise pulses, each presented at an A-weighted sound-pressure level of 70 dB and at a fixed 0-deg azimuth. Aural targets (ATs) were separated into three distances ranging from 24 to 170 in. The time from onset (interstimulus onset interval or ISOI) of the stimuli varied from 6 to 150 ms. Results indicate that aural apparent movement is readily source displayed in depth (i.e., delta movement). As in studies with apparent movement in lateral dimensions, the relative timing of stimuli is crucial for the appearance of apparent movement [$f(8,32)=8.16$, $p<0.001$]. The relative distance between speakers was also found to be a

significant factor for the appearance of apparent movement [$f(16,64) = 3.66$, $p < 0.001$]. These findings will be discussed further as they relate to previous research conducted on motion in the auditory system.

10:45

3aSP10. Apparent movement of sounds originating in different hemifields. Micah Green, Jose A. Gallegos, David R. Perrott, and Nick Lin (Dept. of Psych., California State Univ., 5151 State University Dr., Los Angeles, CA 90032)

In the current study, five participants were asked to describe the sequence of two sounds that originated from the front and back hemifield, relative to the subject. The stimuli were 50-ms broadband noise pulses presented at an A-weighted sound-pressure level of 70 dB. Across sessions, the distance between speakers varied from 78 to 234 in. Within sessions, the delay between the onset (interstimulus onset interval or ISOI) of the first and second speakers varied from 4 to 150 ms. Both ISOI and inter-speaker distance affected the participant performance on the task [$f(8,32) = 8.21$, $p < 0.001$]. It is clear from these results that continuous movement originating from the front and back hemifields [$f(2,8) = 14.06$, $p < 0.001$] occurs over the same time intervals as on the horizontal azimuth. Implications for motion research will also discussed.

11:00

3aSP11. Auditory motion aftereffect and its position-transfer characteristics. C. J. Dong, N. V. Swindale, P. Zakarauskas (Dept. of Ophthalmology, Univ. of British Columbia, 2550 Willow St., Vancouver, BC V5Z 3N9, Canada), V. Haywood (McGill Univ., Montreal, Canada), and M. S. Cynader (Univ. of British Columbia, Vancouver, BC V5Z 3N9, Canada)

To date, there have been only a few studies of the auditory motion aftereffect (aMAE). All of these studies used *simulated* moving sound as adapting or test stimuli. In the present study, *real* moving sound was used as both the adapting and test stimuli. The real moving sound was generated with a loudspeaker mounted on a robot arm which was able to move smoothly and quietly in three-dimensional space. Six subjects were tested in an anechoic chamber with white noise as both the adapting and test stimuli. Results from experiment 1 showed a clear aMAE in all subjects involved. The magnitude of the aMAE tended to increase up to the highest velocity tested (<30°/s). In experiment 2, the adapting and test stimuli were presented across two separate spatial ranges, and the magnitude of the aMAE was measured as a function of the distance between the two separate loci. Results show that when the adapting and test ranges are 20° apart, the aMAE is weak. However, when the two ranges abut, there is a noticeable aMAE, although its magnitude is smaller compared to that when adapting and test ranges overlap. The detailed space constant is under investigation.

11:15

3aSP12. Estimating the spectrum of a moving sound source. M. F. Neelon and Rick L. Jenison (Dept. of Psych., Univ. of Wisconsin, 1202 W. Johnson St., Madison, WI 53706)

A recent analysis of moving sound sources has revealed that information in the dynamic acoustic array can specify fundamental attributes of the source, such as its distance, speed, and spectrum [R. L. Jenison, Ecol. Psych. (1997)]. A model based upon the Kalman filter is introduced which successfully estimates such distal properties using only observed proximal information. The experiment investigated whether 3 specific forms of acoustic information (amplitude, frequency, interaural phase differences) are necessary and sufficient to estimate the intrinsic frequency of a simulated moving sound source. Simulations show the model most accurately estimates intrinsic frequency when all three sources of acoustic information are included in the signal; when dynamic changes in one or more of these parameters are removed, the model's performance degrades. To test human subjects, a technique was developed to generate moving signals from a sample of a natural moving sound source. This process allowed for precise control over the signal's virtual and acoustic parameters. Subjects in a 2AFC design are required to guess the unshifted spectra of the moving stimuli as dynamic changes in the acoustic parameters are systematically removed. [Work supported by NSF.]

WEDNESDAY MORNING, 3 DECEMBER 1997 CALIFORNIA ROOM, 8:00 TO 11:55 A.M.

Session 3aUW

Underwater Acoustics and Acoustical Oceanography: Direct and Inverse Methods in Strongly Range-Dependent Environments I

B. Edward McDonald, Chair
NATO SACLANT Centre, CMR 426, APO, AE 09613

Chair's Introduction—8:00

Invited Papers

8:05

3aUW1. Multipath compensation in shallow-water environments using a virtual receiver. Martin Siderius (Saclant Undersea Res. Ctr., Viale San Bartolomeo 400, 19138 La Spezia, Italy, martin@saclantc.nato.int) and Darrell R. Jackson (Univ. of Washington, Seattle, WA 98105)

An acoustic technique for the compensation of signal distortion due to propagation is developed in analogy with an astronomical technique in which light from a bright ''guide star'' is used to correct atmospheric aberration of weaker objects that are nearby in the angular sense. The acoustic technique investigated here uses a vertical array to receive both the signal from a broadband ''guide source'' and an unknown ''objective source'' which propagates over a partially shared path. The algorithm is a spatial-temporal cross correlation of the two signals and is termed a ''virtual receiver,'' as the output approximates the signal that the unknown source would

produce at the location of the guide source. By strategically locating the guide source, many of the distorting effects of an unknown propagation region can be removed, including mode coupling and multipaths. Results are given for a variety of numerical experiments in both the time and frequency domains. The numerical simulations are used to illustrate the virtual receiver algorithm in range-dependent, shallow-water environments exhibiting lossy propagation and mode coupling.

8:25

3aUW2. Internal wave strength inversion and localization using long-range acoustic signals. Kevin D. Heaney[a)] (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093)

The ATOC95 experiment was conducted in the winter of 1995 with transmission paths from Pioneer Seamount to Hawaii (range = 3515 km). The internal wave environment is strongly range dependent, and significant mode coupling occurs to the late arriving energy. The ability to localize a source is an indication of the remaining source information in the signal. A mode travel time difference range estimator was developed and shown to be biased by internal wave-induced mode coupling. A range estimator was developed which matched the received wavefronts with those of predicted wavefronts at various ranges. Forward simulations were done using a frequency interpolation method of the PE, and a set of random realizations of the Garrett–Munk spectra for internal waves. With this ensemble of calculations as a forward model, three different measures were calculated from the data and simulations to do an inversion for the background internal wave strength in the ocean. A method was developed to simultaneously invert the source location and the internal wave strength using the mode arrival time and mode spread as the two observables. [a)]Present address: Sci. Applications Intl. Corp., 10260 Campus Pt. Dr., San Diego, CA 92121.

8:45

3aUW3. Coupled wave-number integration approaches to range-dependent, seismoacoustic modeling. Henrik Schmidt (SACLANT Undersea Res. Ctr., Viale San Bartolomeo 400, I-19138 La Spezia, Italy and MIT, Cambridge, MA 02139, henrik@keel.mit.edu)

Wave-number integration (WI) approaches directly implement the ''exact'' spectral formulation for the seismoacoustic field in ocean waveguides. Therefore, these modeling approaches are inherently limited to range-independent, or horizontally stratified, environments. However, WI possesses a number of unique seismic capabilities, and various wave-field coupling approaches have recently been combined with WI to extend its applicability to two-way, seismoacoustic propagation in strongly range-dependent environments. These hybrid modeling approaches use a stepwise discretization of the environment similar to the one used in coupled-mode models. The spectral super-element approach [J. T. Goh and H. Schmidt, J. Acoust. Soc. Am. **100**, 1409–1420 (1996)] then uses a Galerkin boundary-element formulation for the coupling between the range-independent sectors (super elements), while the virtual source approach [H. Schmidt, Proc. SWAC-97, Beijing (Apr. 1997)] uses an approximate, but highly efficient, local solution to the coupling problem. The performance of these new approaches has been demonstrated using a new class of two-way seismoacoustic benchmark problems with both weak and strong range dependence. Here, the new modeling approaches will be described, and their unique seismoacoustic capabilities will be discussed and illustrated by application to, e.g., the deep-ocean waveguide coupling of near-shore tomography sources, and analysis of potential seismoacoustic coupling mechanisms for oceanic *T* phases. [Work supported in part by the Office of Naval Research.]

9:05

3aUW4. Forward modeling techniques for range-dependent problems. Michael D. Collins, Joseph F. Lingevitch (Naval Res. Lab., Washington, DC 20375, collins@ram.nrl.navy.mil), and William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY 12180)

This paper will discuss progress that has been made in modeling range-dependent propagation problems and outline some of the unresolved problems. The parabolic equation (PE) method has proven to be the most useful approach for solving range-dependent problems. Finite-difference solutions and range-dependent mode and spectral solutions are useful for confirming the accuracy of PE solutions and for solving problems that PE techniques cannot handle (or cannot yet handle). Combinations of these approaches, such as the coupled mode PE, are also useful. In the mid 1980s, it was widely believed that there was not much left to do in propagation modeling. The amount of work that has been done in the past decade or so is evidence of how shortsighted this view was. Although there now exist accurate and efficient techniques for solving a wide variety of range-dependent problems, there still remain important problems to be solved. [Work supported by ONR.]

Contributed Papers

9:25

3aUW5. Yellow Shark '95: Experimental and inversion results for broadband acoustic propagation over a soft clay bottom. Jean-Pierre Hermand and Martin Siderius (SACLANT Undersea Res. Ctr., Viale San Bartolomeo 400, 19138 La Spezia, Italy, martin@saclantc.nato.int)

In May 1995, Saclant Centre performed broadband (200–800 Hz) acoustic measurements near Elba Island off the coast of Italy as part of the Yellow Shark '95 experiments. The experimental site has a soft clay layer bottom that varies in thickness from 3 to 10 m and has a sound speed less than the water column. The data presented here are from a range-dependent transect where the water depth varies from 67 m at the source location to approximately 120 m at the vertical arrays located every 8 km from 8 to 40 km from the source. The vertical arrays were sparsely populated with just four hydrophones positioned throughout the water column. A modal analysis is given to explain the frequency bands where high transmission loss is observed in the data. The observed data are inverted by matching the fields with modeling results. That is, using range-dependent modeling, the bottom type was varied in a search to fit the large-scale features of the modeled and observed transmission loss data. The results show that large-scale features can be used to identify some properties of the bottom such as the existence and thickness of the slow clay layer.

9:40

3aUW6. Measurement and analysis of the propagation of sound from the continental slope to the continental shelf. Ching-Sang Chiu, Kevin B. Smith (Naval Postgraduate School, Monterey, CA 93943), James F. Lynch, Glen G. Gawarkiewicz, Robert S. Pickart, Brian Sperry (Woods Hole Oceanograph. Inst., Woods Hole, MA), James H. Miller (Univ. of Rhode Island, Narragansett, RI), and Allan R. Robinson (Harvard Univ., Cambridge, MA)

Carried out jointly by the Woods Hole Oceanographic Institution (WHOI), Naval Postgraduate School (NPS), Harvard University, and the University of Rhode Island (URI), the measurement program of an integrated acoustic-oceanographic field study called Shelfbreak PRIMER took place in the Middle Atlantic Bight. One of the goals of Shelfbreak PRIMER is to characterize and understand the propagation of sound from the continental slope to the continental shelf, including the effects of shelfbreak frontal processes, seasonal stratification, and topographic variations. The field work included two intensive 3-week experiments, one in July 1996 (summer) and the other one in February 1997 (winter). In particular, each of the two experiments employed a suite of acoustic and oceanographic sensors including several transceivers/sources and two vertical hydrophone arrays (VLAs) straddling the shelfbreak front and a SeaSoar that provided several high-resolution, three-dimensional surveys of the frontal region. The results from a data and modeling analysis of the pulse transmissions from a fixed sound source on the slope to a fixed VLA on the shelf are discussed. The objective of this analysis is to gain fundamental insights into the tidal-to-seasonal variability of the amplitudes and travel times of the strongly coupled normal modes. [Work supported by ONR.]

9:55–10:10 Break

10:10

3aUW7. Three-dimensional propagation effects near the Mid-Atlantic Bight. Kevin B. Smith (Dept. of Phys., Naval Postgraduate School, Monterey, CA 93943), Ching-Sang Chiu (Naval Postgraduate School, Monterey, CA 93943), James H. Miller (Univ. of Rhode Island, Narragansett, RI 02882), James F. Lynch, and Glen G. Gawarkiewicz (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

A recent set of experiments sponsored by the Office of Naval Research has been conducted in the vicinity of the Mid-Atlantic Bight shelf break region off the coast of New Jersey. The environment in this region is highly variable in both range and bearing due to the shelf break slope and its associated front. One of the goals of this Shelf Break Primer project is a determination of the influence of azimuthal coupling on the 3-D propagation. By using a fully 3-D PE model and comparing with results from the associated 2-DxN PE approximation, a quantitative assessment of the influence of azimuthal coupling can be made. Some sample benchmark cases are performed to illustrate this method. Using this approach, the significance of both the sloping bathymetry and the shelf break front will be considered separately and combined. In addition, the influence on the arrival time structure of a broadband pulse, similar to a signal employed during the experiments, will be determined. Characteristic features of such 3-D effects in the pulse propagation will then be explored in the available data sets. [This work is supported by the Office of Naval Research, Code 321OA.]

10:25

3aUW8. Source localization bias due to forward model uncertainties in matched-field processing. N. Ross Chapman and Ronald T. Kessel (School of Earth and Ocean Sci., Univ. of Victoria, P.O. Box 3055, Victoria, BC V8W 3P6, Canada)

Matched-field source localization (MFSL) in range-dependent media requires fast realistic forward models of sound propagation. No model replicates the real environment in all its detail, however, which effectively means that the source is being localized in the wrong environment, at a position that may not agree with the actual. As a rule, the agreement improves as the model is made more realistic, but this usually increases the computation time and complexity of the model, especially in range-dependent environments. MFSL must, therefore, balance accuracy and speed against model realism. The objective here is to demonstrate and account for consistent errors in the estimated range and depth of a source when the forward model differs from the actual environment. Range bias is demonstrated for data collected in the Pacific Shelf sea trials, for example, when bathymetric variation was at first ignored for speed and simplicity in the MFSL analysis, and then taken into account using a simple bias rule. Range and depth biases due to incorrectly modeled sediment properties and bathymetry in range-dependent media are explored more generally using the theory of adiabatic modes and numerical simulations. [Work supported by ONR and Department of National Defence, Canada.]

10:40

3aUW9. Sound propagation in regions overlaying gaseous range-dependent sublayers. Natalia A. Sidorovskaia (Phys. Dept., Univ. of New Orleans, New Orleans, LA 70148) and Michael F. Werby (Code 7181, NRL, Stennis Space Center, MS 39529)

Some regions of the ocean consist of bottom layers, which have imbedded bubbles. The bubbles may consist of methane gas. The consequence is that the particular layer has a sound speed, which is lower than in the overlying layer. Depending on the wavelength of the sound source relative to the layer thickness, sound may either reflect off the layer, or it may partially tunnel through the layer to the next set of layers. This effect is frequency dependent, and at some point, when the wavelength in the layer is small relative to the layer thickness, no sound will traverse the layer. The effect of this is that the modal eigenspectrum will be altered and will be largely characterized by the gaseous layer structure in horizontal and vertical directions. The influence of range-dependent gaseous regions on modal spectrum and measured signal structure in different underwater waveguides, which can be considered as strongly range dependent for spatially localized gaseous regions, is modeled by the normal mode code SWAMP for different frequency ranges. [Work sponsored by ONR, the Naval Research Laboratory, and the University of New Orleans.]

10:55

3aUW10. Two-dimensional acoustic field measurements in a complex, shallow-water waveguide. Laurence N. Connor, George V. Frisk, Kyle M. Becker, and Cynthia J. Sellers (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

A shallow-water acoustics modal mapping experiment (MOMAX) was conducted in March 1997 in the vicinity of the East Coast STRATAFORM site. An acoustic modal field was established in a shallow-water environment using fixed and drifting sources radiating several pure tones in the range 50–300 Hz. The field was then densely sampled on a horizontal plane at midwater depth using freely drifting buoys equipped with suspended hydrophones and a radio telemetric link to the ship for data acquisition. GPS and acoustic navigation were used on each buoy for precise spatial tracking to a maximum range of 10 km. An overview of the experiment is presented emphasizing buoy configuration and drift paths, signal processing, and acoustic field mapping. Specific buoy deployments will be examined with regard to the lateral variability of the acoustic field using data obtained from the two-dimensional buoy drift tracks. [Work supported by ONR.]

11:10

3aUW11. Modal evolution in a three-dimensionally varying shallow-water waveguide. Kyle M. Becker, George V. Frisk, and Laurence N. Connor (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543, kbecker@whoi.edu)

Theory shows that the wave-number spectrum of a propagating normal-mode field is a function of position for a complex, shallow-water waveguide whose acoustic properties vary in three spatial dimensions. To demonstrate these effects, a modal mapping experiment (MOMAX) was conducted in the vicinity of the East Coast STRATAFORM site. Using

both fixed and drifting source configurations, several pure tones in the range 50–300 Hz were transmitted to a field of freely drifting buoys, each containing a hydrophone, GPS and acoustic navigation, and radio telemetry. The resulting two-dimensional, synthetic aperture planar array can then be used to determine the evolution of the normal modes with position. By measuring the spatially varying spectral content of the modal field, propagation characteristics of the waveguide may then be inferred. Preliminary results will be presented for different source and buoy configurations demonstrating modal evolution as a function of position. Supporting environmental and model data will also be presented and correlated with the experimental results. [Work supported by ONR.]

11:25

3aUW12. Submeter range determination in an ocean environment using differential moving-source kinematic GPS software. James A. Doutt, Keith von der Heydt, Calvert Eck, Laurence N. Connor (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543, jdoutt@whoi.edu), Darren Cosandier, and Hugh Martell (Waypoint Consulting, Inc., Calgary, AB T2C 2X5, Canada)

A relative accuracy of about 1 m in range between a ship-towed sound source and drifting buoys was required for the analysis of data from the MOMAX shallow-water acoustics experiment performed off the New Jersey coast in March 1997. The maximum separation between the ship and the buoys was less than 10 km. Single-frequency GPS receivers providing both code and carrier phase information were installed on the ship and the buoys, and their outputs were recorded for postprocessing at sea. Differential moving-source kinematic software was used to determine the ranges between the ship and the buoys to submeter accuracy. [Work supported by ONR.]

11:40

3aUW13. Predictions of ambient noise directionality and array response in range-dependent environments with CANARD and RANDI-2. Richard J. Brind (Defence Evaluation and Res. Agency, Sonar Systems Dept., Winfrith Technol. Ctr., Dorchester, Dorset DT2 8XJ, United Kingdom, rbrind@dra.hmg.gb), Chris H. Harrison, and Rachel M. Hamson (BAeSEMA, Surrey KT3 4LH, United Kingdom)

There has been recent interest in the prediction of array performance in shallow-water environments. Several models have been developed in the UK over the last few years for predicting ambient noise directionality and array response in complex range-dependent environments. One model, CANARD, has been developed to provide quick and reliable predictions in realistic environments. It is a ray-based model but avoids using a propagation model in the usual sense. Environmental information can be read in from databases or specified manually. The model can treat nonuniform distributions of noise sources, wide vertical separation of hydrophones, cross slopes, and near-surface and duct leakage effects. Another model, RANDI-2, is a full-wave model based on the SUPERSNAP propagation model. It has been improved to calculate the wind noise component using virtual modes rather than the previous numerically intensive fast-field program method, and to improve accuracy at short range for the calculation of source and near-field shipping contributions. Software to generate input parameters from databases has also been produced. Details of the theory and capabilities of the models are explained. Examples of predictions in the SW approaches for a number of different types of array are presented. [Work supported by MOD(UK).]

WEDNESDAY MORNING, 3 DECEMBER 1997 ADOBE ROOM, 9:00 A.M. TO 12:00 NOON

Meeting of Accredited Standards Committee S2 on 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 (and Subcommittees ISO/TC 108/SC1, SC2, SC3, SC5, and SC6)

D. J. Evans, Chair S2 and Chair of the U. S. Technical Advisory Group (TAG) for ISO/TC 108, Mechanical Vibration and Shock

National Institute of Standards and Technology, Acoustics, Mass and Vibrations Group, Building 233, Room A147, Gaithersburg, Maryland 20899

Accredited Standards Committee S2 on Mechanical Vibration and Shock. Working group chairs will present reports of their recent progress on writing and processing various shock and vibration standards. There will be a report on the interface of S2 activities with those of ISO/TC 108 (the Technical Advisory Group for ISO/TC 108 consists of members of S2, S3, and other persons not necessarily members of those committees), including a report on the meeting of ISO/TC 108 with Subcommittees SC1 and SC5 (11-19 September 1997 in Berlin, Germany), ISO/TC 108/SC4 (1-5 September 1997 in London, U.K.), ISO/TC 108/SC3 (October 1997 in San Juan Capistrano, California) and ISO/TC 108/SC6 (7-11 September in St. Petersburg, Russia).

Scope of S2. Standards, specifications, methods of measurement and test, and terminology in the 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 3pEA

Engineering Acoustics: General Topics

Kim C. Benjamin, Chair
Nabal Undersea Warfare Center, Newport, Rhode Island 02841

Contributed Papers

1:00

3pEA1. High-definition temporal/frequency analysis of sampled data sets. George J. Frye (Frye Electron., Inc., PO Box 23391, Tigard, OR 97281, gfrye@frye.com)

The FFT is a successful steady-state analysis tool, but falls short when temporal analysis is needed. A class of filters designed around a Gaussian windowed DFT is described that, when applied to signal data sets in a way similar to that of a wavelet analysis, yields high-definition, fractional octave bandwidth, temporal analysis. The filter set's desired properties are temporal resolution that is inversely proportional to bandwidth and directly proportional to frequency, no assumption of signal periodicity, and the trade-off of accuracy against temporal window width. The main undesirable property is the computation needed to perform an analysis. Examples are chosen to show the need of a good display technique. They include analysis of a mathematical impulse, a hearing aid output as driven by a transient tone burst, and a room reflection with a broadband chirp signal. The author wishes to acknowledge Mike Day and Ilker Gungor of Frye Electronics for help in basic concepts and software programming.

1:15

3pEA2. Using photoacoustic signals to localize small gas leaks. Serdar H. Yönak and David R. Dowling (Dept. of Mech. Eng. and Appl. Mech., Univ. of Michigan, Ann Arbor, MI 48109-2121, serdar@engin.umich.edu)

In the manufacture of devices that are to be filled with pressurized gases or liquids, efficient localization of leaks is a critical quality control process. Current assembly-line leak localization techniques such as bubble visualization, sniffing devices, or helium-mass spectrometers are either subjective or time consuming. Photoacoustics has already been used to detect small leaks but has not been extended to the localization of these leaks. This work presents a quick and objective leak localization method. Sulfur hexafluoride is used as the tracer gas. Photoacoustic sound is generated by rapidly scanning a carbon dioxide laser tuned to 10.6 μm over a suspected leak area. The sound is measured by multiple microphones. Using the microphone output and signal processing techniques, the location of the leak is determined. Leak location uncertainty due to microphone placement, laser scanning, and leak size is discussed. A signal-to-noise-ratio-dependent leak detection limit is established. Experimental results are presented. [Work supported by Ford Motor Company.]

1:30

3pEA3. Sound source segregation using the inter-channel differences in both intensity and phase. Mariko Aoki and Shigeaki Aoki (NTT Human Interface Labs., 1-1 Hikarinooka Yokosuka, Kanagawa, 239, Japan, mariko@nttspch.hil.ntt.co.jp)

A method is proposed for segregating sounds from multiple sources recorded on a two-channel system. It uses the differences in both intensity and phase, between the channels. In addition, the relation between frequency resolution and sound quality of the proposed method is examined. In the tests, three pairs of mixed sounds were used, namely, male speech with female speech, two kinds of female speech (the same person with different phrases), and a cock's call with female speech. Frequency resolution was varied between 5 and 80 Hz. In the subjective test, subjects listened to the original sound plus sounds segregated at five different resolutions: 5, 10, 20, 40, and 80 Hz. Then they estimated the quality of these sounds on a five-point scale. Moreover, the correlation coefficient between the original sound and each segregated sound is discussed. In the case of a human voice, it is determined that the optimal frequency resolution is 10 or 20 Hz. The interesting fact is that the optimal frequency resolution was not the highest. Compared with the human voice, lower frequency resolution may be enough to segregate a cock's call.

1:45

3pEA4. Sound absorption by nonhomogeneous thin porous layers. Noureddine Atalla, Raymond Panneton (Dept. Mech. Eng., Universite de Sherbrooke, QC, Canada), and Jean-Franois Allard (Universite du Maine, Le Mans, France)

Sound absorption and sound transmission of nonhomogeneous porous materials made of random or regular arrays of porous media is investigated using the finite-element method. The configuration studied consists of a column of porous materials inserted in an infinite rectangular waveguide. In the finite-element model, an equivalent fluid model for the porous material is coupled to the modal behavior in the waveguide. By comparison with analytical calculations, the difference between absorption mechanisms in three-dimensional porous patchworks and simpler stratified materials is investigated. In particular, the importance of nonpropagative mechanisms is enlightened. Also, the transmission loss performance of nonhomogeneous porous materials compared to homogeneous porous material is discussed. [Work supported by Bombardier, Inc., Canadair and NSERC.]

2:00

3pEA5. Transmission loss measurement in time domain using an impulse technique in ducts. Andre L. Cherman and Roberto A. Tenenbaum (Federal Univ. of Rio de Janeiro, CT bloco G sala 204, CP 68501, Ilha do Fundao, Rio de Janeiro, RJ 21945-970, Brazil, cherman@serv.com.ufrj.br)

The method proposed is an option to the one described by ISO-140, which needs two reverberant chambers. It consists of a duct 3 m long with a cutoff frequency around 1300 Hz, with a loudspeaker at one end and an anechoic termination at the other. The sample is placed 1 m from the loudspeaker and two microphones are placed around 500 mm from the sample, the first between the loudspeaker and the sample and the second between the sample and the anechoic termination. An impulse is generated and captured at microphone 1 as the incident pulse at the sample and at microphone 2 as the transmitted pulse through the sample. The impulse size is controlled in order to have enough energy in the frequency range desired and not to be too long to mix with the pulse reflected by the sample. In the frequency domain, the energy difference between the signals is compared with the mass law. The ultimate goal of this research is achieving a simple technique for evaluation of transmission loss in multilayered partitions.

2:15

3pEA6. Sound energy distribution at an intersection of underground tunnels: Full scale experiments. Hiroyuki Imaizumi, Michiaki Kinoshita, Sunao Kunimatsu, and Takehiro Isei (Safety Eng. Dept., Natl. Inst. for Resources and Environ., 16-3 Onogawa Tsukuba, 305 Japan)

Experiments on sound propagation in small-scale tunnels with a branch have been carried out from the viewpoint of safe and rapid information transmission underground. The influences of intersection angles and acoustical properties of inner walls of the tunnels on sound propagation were examined in previous studies [J. Acoust. Soc. Am. **100**, 2567 (1996)]. In the present work, sound propagation in a full-scale underground tunnel with a branch is considered. The experimental tunnels are supported by iron frames and concrete lagging, and the cross sections are arch shaped with an area of 4.2 m^2. Experimental results indicate influences of the intersection on propagation attenuation at a frequency of around 200 Hz. Whole tendencies of sound attenuation are also affected by relative arrangements of sound source and receivers, and sound attenuates more at the receivers, which are out of sight from the source than for those in sight. For the purpose of the sound-field estimations, the characteristics of sound propagation are examined by numerical calculations using both an analogy with a resonant silencer with a branch and an image source model at the receivers, which are out of sight. Results of the numerical calculations relatively agree with the experimental results.

2:30

3pEA7. On the modeling of sound propagation through capillary tubes with mean flow. Jeong-Guon Ih (Dept. of Mech. Eng., Korea Adv. Inst. of Sci. and Technol., Science Town, Taejon 305-701, Korea)

In order to predict the transmission loss of catalytic converters in automotive exhaust systems, efforts have been devoted recently to wave propagation through a capillary tube with a steady flow. Several theoretical models also exist. In this paper, general linearized governing equations are discussed and the characteristic features of each model are shown in conjunction with methods for solving the equations. Characteristics of the forward and backward propagating acoustic waves are investigated under the influence of steady flow and the dependency on the shear wave numbers is checked. Additionally, the presence of hydrodynamic wave components is examined under the steady flow condition. The application limits of the concurrent acoustic models are discussed as the shear wave number varies. Several prediction methods are applied to the calculation of the transmission loss of a stack of monoliths with and without mean flow considering the effects of radial components.

WEDNESDAY AFTERNOON, 3 DECEMBER 1997 FRIARS/PADRE ROOMS, 1:30 TO 2:35 P.M.

Session 3pID

Interdisciplinary: Hot Topics in Acoustics

Kenneth J. Plotkin, Chair
Wyle Laboratories, 2001 Jefferson Davis Highway, Suite 701, Arlington, Virginia 22202

Chair's Introduction—1:30

Invited Papers

1:35

3pID1. Conversational speech recognition. Thomas H. Crystal (IDA Ctr. for Commun. Res., Thanet Rd., Princeton, NJ 08540, crystal@ccr-p.ida.org)

A hot topic in speech recognition is developing technology for the automatic transcription of telephone conversations. The recognizer must contain robust language, pronunciation, and acoustic models that embody the world and topic knowledge and the understanding of syntax and pronunciation, which the talkers have and use in decoding each other's acoustic signals. Partly because of the talkers' shared knowledge and the casual, unprepared nature of the speech, the signals have dysfluencies, incomplete and ungrammatical expressions, and ''lazy,'' reduced articulation of words. Conversational speech recognition error rates, measured in the NIST Hub-5 evaluations, are 45% for English and 66% to 75% for Spanish, Mandarin, and Arabic. To improve this performance, the shared knowledge must be represented in a mathematical framework, which facilitates the efficient search of the sentences of a language to decode the speech. Recent work, including workshops at Rutgers CAIP and Johns Hopkins CLSP, has included the investigation of, among other techniques, multistream processing, frequency warping, adaptation of pronunciation and acoustic models of phones, pronunciation modeling, syllable-based recognition, dysfluency and discourse-state language models, and link grammar parsing. This talk will review how knowledge is represented in the recognizer architecture, searching procedures, and the results of the various investigations.

1:55

3pID2. Near-field acoustical holography—interior problems. Earl G. Williams (Physical Acoust., Naval Res. Lab., Washington, DC 20375)

Near-field acoustical holography (NAH) provides a practical solution to a challenging inverse problem: Calculation of the normal velocity on the surface of a vibrator from the pressure field measured in the acoustic near field of the vibrating surface. Furthermore, NAH uses this same pressure measurement to compute the pressure field at the surface, and throughout the volume of interest. Thus structural vibration and acoustic radiation are reconstructed from a single pressure-array measurement. NAH has been used extensively throughout the years to study the vibration, radiation, and scattering applied to exterior problems, in particular, plates and cylinders. Recent applications, however, have been for interior problems in which the vibrations of the walls surrounding a cavity, as well as the pressure field throughout the cavity, are determined from a two-dimensional, conformal array interior pressure measure-

ment. The aircraft fuselage is a good example of recent applications. Recent research is poised to solve the inverse problem for partially correlated noise sources. This research centers on a spherical measurement antenna for the interior problem, with the inversion relying on decomposition into spherical harmonics along with very fast numerical quadrature algorithms. Most current in this approach is the use of analytic continuation to translate the measured field to a different origin, without the need for further measurement.

2:15

3pID3. Room criteria. Richard J. Peppin (Larson-Davis, 5012 Macon Rd., Rockville, MD 20852)

This paper presents the status of standards dealing with this very contentious subject. Room criteria are contentious because, for years, the acoustical community has been trying, with minimal success, to come to some consensus on the types of descriptors, measurement procedures, and rating schemes based on steady state noises in rooms. (What constitutes steady state is not agreed upon.) Evidence of this long standing problem is the new ANSI standard S12.2-1995: Even though written after perhaps 10 years effort, it is still not satisfactory to many, if not most, people. This presentation tries to provide a perspective on this issue, not from an expert's point of view, but from someone immersed in the controversy for a long time. The issues that need to be addressed, the current and proposed methods, and the problem with data collection, funding, and research will be discussed. Audience input and reaction will be solicited if time permits.

WEDNESDAY AFTERNOON, 3 DECEMBER 1997 SIERRA ROOM, 1:00 TO 2:50 P.M.

Session 3pNS

Noise, Musical Acoustics and Underwater Acoustics: Session in Honor of Robert W. Young

Richard M. Guernsey, Cochair
R. M. Guernsey and Associates, P.O. Box 1517, Morristown, New Jersey 07962-1517

Tony F. W. Embleton, Cochair
80 Sheardown Drive, Box 786, Nobleton, Ontario L0G 1N0, Canada

Chair's Introduction—1:00

Invited Papers

1:05

3pNS1. Robert W. Young's eminence in musical acoustics. Daniel W. Martin (7349 Clough Pike, Cincinnati, OH 45244)

Interest in musical sound led Robert W. Young into physics graduate study at the University of Washington, where his 1934 doctoral thesis reported on standing waves within the Boehm flute. During subsequent research at C. G. Conn Ltd., Young and co-inventor Hugo Schuck developed the Conn Chromatic Stroboscope. It was the most precise frequency measuring instrument for acoustical research in the audio frequency range. Young's book of tables for instrument users, *From Frequency to Cents* [1939; 2nd ed. (1952)], related frequency to cents deviation from the equally tempered musical scale. Research in musical acoustics, musical instruments, and pitch perception was especially benefited for decades, before computers became available. Young's 63 research publications in musical acoustics concerned tuning of wind instruments, vibration of piano strings, temperature and humidity effects upon timing, inharmonicity of string vibration modes, and the resulting stretched scale of piano tuning. He was the associate editor for musical acoustics for the *Journal of the Acoustical Society of America* for 45 years, exerting a positive, and often creative, influence on research by authors around the world. He also served as associate editor for the acoustical patent review section of this *Journal* from 1943 to 1977, and personally reviewed thousands of patents, many on musical sound.

1:20

3pNS2. Robert Young and the measurement of community noise. William J. Galloway (Tarzana, CA 91356)

Throughout his career in acoustics Robert Young has maintained a major interest in the development of good measurement practices for a variety of acoustical measurements. One of his major interests has been in the measurement of outdoor sounds as experienced by community residents. He has been one of the long-term, principal advocates of the use of an A-weighted sound level as a primary measure for noise as it relates to the response of people. He has shown that the A-weighted sound level can be used to describe the acoustical environments produced by sound sources as diverse as those of the quasisteady noise at an industrial site, the intermittent noise such as that produced by sequences of aircraft operations around an airport, to the impulsive sound produced by a sonic boom from the landing of a space shuttle orbiter.

1:35

3pNS3. Robert W. Young's contributions to acoustical standards. Alan H. Marsh (5092 Tasman Dr., Huntington Beach, CA 92649)

The citation for the award to Robert W. Young of Honorary Fellowship in the Acoustical Society of America, and the accompanying encomium, mention Bob's contributions to the development of national and international standards. The contributions were many, sustained, and important. Over a span of five decades since the 1940s, his contributions to acoustical standards cover a range of acoustical instruments; measurement and evaluation of sound fields in rooms, outdoors, and in the ocean; and acoustical terminology. The quality of acoustical standards produced through the Acoustical Society of America, the International Electrotechnical Commission, and the International Organization for Standardization has been enhanced through the efforts of Robert Young.

1:50

3pNS4. Robert Young and acoustics during the War. Robert S. Gales (1645 Los Altos Rd., San Diego, CA 92109)

This paper will discuss Bob Young's contributions to the field of acoustics through his work in San Diego, CA during World War II. The contributions will be highlighted by anecdotes from personal experiences.

Contributed Papers

2:05

3pNS5. Algorithm for BEM calculations to simulate sonic boom noise penetration into the ocean. Victor W. Sparrow (Graduate Program in Acoust., Penn State Univ., 157 Hammond Bldg., University Park, PA 16802, sparrow@helmholtz.acs.psu.edu)

Robert W. Young has made many contributions to our understanding of how sound can get transmitted from air to water. One important contribution is Bob's measurements of sonic boom noise penetration into the ocean from supersonic airplanes [J. Acoust. Soc. Am. **44**, 392 (1968)]. Bob's last words in this paper were ''I would voice the hope that this experimental evidence will inspire some further theorizing to explain the observations.'' Now almost 30 years after this statement, some of us are still theorizing. The present paper gives a quick overview of the state of knowledge in sonic boom penetration into the ocean and introduces a new boundary element method (BEM) approach to solving the problem of evanescent sound penetrating the ocean surface. The method is based on a mesh made from discretizing the surface of the air–water interface. Perhaps along with analytical formulations, finite difference simulations, and other experimental observations already available for determining the sonic boom noise penetration into the ocean, such new BEM simulations will help us answer the questions which Bob Young has posed. [Work supported by NASA Langley Research Center.]

2:20

3pNS6. On the units of Sabine absorption and the Sabine absorption coefficient. J. Pope (Pope Engineering Co., P.O. Box 236, Newton Centre, MA 02159)

A century ago, W. C. Sabine demonstrated that sound absorption is fundamental to room acoustics, and introduced the idea for a measure of absorption that has become known as Sabine absorption. Because Sabine absorption and area have the same dimension, e.g., square-meters, many acousticians presume that Sabine absorption has the units of area. For a number of years Robert W. Young has been trying to convince us otherwise, arguing that a dimensionless logarithmic unit distinguishes Sabine absorption from area. [By dimensionless logarithmic unit is meant a scale factor that defines unity on a log scale. The neper (Np) and the decibel are two such quantities.] The metric sabin (Sa) is a unit of Sabine absorption. Defining the metric sabin to be one-half neper square-meter (1 Sa=0.5 Np m^2) introduces a logarithmic unit while preserving the numerical scaling of current measurement practice. It follows that the unit of the Sabine absorption coefficient is the half-neper, and a difference between the Sabine absorption coefficient and coefficients from other methods becomes explicit. Among other things, a partial explanation for the experimental observation that Sabine absorption coefficients can exceed unity is provided.

2:35

3pNS7. Room noise criteria standards: What features are important? Hsien-sheng (Jason) Pei, Robert D. Hellweg, Jr. (Digital Equipment Corp., 129 Parker St., Maynard, MA 01754), and Richard Peppin (Larson-Davis, Rockville, MD 20852)

At the joint meeting of ASA and NOISE-CON 97 in June 1997, the authors organized a special session on room noise criteria, in which a ''questionnaire'' on the features of several existing noise criteria methods was answered by panel experts and distributed to the noise control community for their inputs. The four methods were: the A-weighted sound-pressure level; the noise criteria (NC) tangency method contained in older versions of the ASHRAE handbook; the balanced noise criterion method (NCB) contained in ANSI S12.2; and the room criteria method (RC) contained in ANSI S12.2 and the current ASHRAE handbook. The features considered were: speech interference, high-frequency annoyance, mid/low frequency annoyance, very low-frequency annoyance, pure tones, temporal fluctuations, user friendliness, wider frequency range, speech privacy, costs, and ability to handle noises with abnormal characteristics and complexity. The questionnaire rating was twofold: (1) how well does each method adequately account for each of the ''features'' (or characteristics) relative to the other methods, and (2) what is the relative importance of each of the ''features'' in a noise criteria method. This paper presents the results of the questionnaires and reports progress since the joint meeting in ANSI S12 WG18 and ASHRAE TC2.6.

Session 3pUW

Underwater Acoustics and Acoustical Oceanography: Direct and Inverse Methods in Strongly Range-Dependent Environments II

B. Edward McDonald, Chair
NATO SACLANT Centre, CMR 426, APO, AE 09613

Contributed Papers

1:00

3pUW1. Three-dimensional, energy-conserving, and reciprocal one-way acoustic wave equations. Oleg A. Godin (School of Earth and Ocean Sci., Univ. of Victoria, P.O. Box 1700, Victoria, BC V8W 2Y2, Canada)

Within the parabolic approximation, preserving energy conservation and reciprocity intrinsic to acoustic fields in stationary media is known to be vital for accurate amplitude predictions in shallow-water environments and for modeling subtle nonreciprocal acoustic effects due to ocean currents. Recently, a wide-angle parabolic equation (PE) has been derived that ensures energy conservation and reciprocity of its solutions in fluids with range-dependent, piecewise-continuous sound speed and density [O. A. Godin, J. Acoust. Soc. Am. **100**, 2835(A) (1996)]. In this paper, the theory is extended to 3-D problems having considerable azimuth coupling. Moreover, a technique is proposed to derive energy-conserving and reciprocal PEs with arbitrary high wide-angle capability. Formulations in cylindrical and Cartesian coordinates are compared. The ability of the 3-D energy-conserving PE to account for cross-range inhomogeneities and horizontal refraction is discussed. The asymptotic accuracy of the energy-conserving parabolic approximation is addressed and compared to that of other paraxial approximations. It is concluded that energy conservation, reciprocity, and an improved description of mode coupling can be achieved simultaneously within the 3-D parabolic approximation without adversely affecting phase accuracy. [Work supported by NSERC.]

1:15

3pUW2. Acoustic diagnostic for the polar front. E. C. Shang and Y. Y. Wang (CIRES, Univ. of Colorado, NOAA, Environ. Technol. Lab., 325 Broadway, Boulder, CO 80303)

The Polar frontal zone is known to have the highest contrast between properties of the divided water masses in Nordic seas. Acoustically, it is a strongly range-dependent environment. Numerical simulations of acoustic propagation in the frequency range of 20–200 Hz have been conducted for typical cases by using the IFD-PE and KRAKEN codes. The adiabaticity is strongly frequency dependent. For lower frequency, the inverted results of the frontal parameters based on the adiabatic modal travel time and modal horizontal refraction data can be obtained analytically using a simple ''few parameter model'' proposed by Kuz'kin [Acoust. Phys. **39**(4), 402 (1993)]. The possibility of estimating the frontal parameters for a more complex realistic model will also be discussed. [Work supported by ONR and NOAA.]

1:30

3pUW3. Coherence of acoustic modes passing through internal solitary waves. James C. Preisig and Timothy F. Duda (Dept. of Appl. Ocean Phys. and Eng., MS 11, Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Packets of internal solitary waves can affect coherence between acoustic modes propagating in shallow water. Shallow water is characterized as having less than 10 or 15 propagating modes, including the influence of an attenuating bottom. Previous work has shown that individual internal waves or packets of waves can scatter low-mode energy into higher modes. Prior results and intuition indicate that the higher modes attenuate more quickly through boundary interaction. The propagation of the internal waves, which couple energy between the modes, can rapidly influence the relative amplitudes and phases of the modes. Numerical simulations are used here to quantify the effects which the relative positions of the source, receiver, and wave packet (or packets) have on the mode coherence. Of particular interest are packets of solitary waves near the source, which can amplify the source. Frequencies of a few hundred Hz and propagation over tens of km are investigated. The results are compared with experimental data. [Work supported by ONR Ocean Acoustics.]

1:45

3pUW4. On energy-conserving boundary conditions for parabolic wave equations. Oleg A. Godin (School of Earth and Ocean Sci., Univ. of Victoria, P.O. Box 1700, Victoria, BC V8W 2Y2, Canada)

Realistic description of sound interaction with the ocean bottom requires environmental models having piecewise-continuous dependence of their parameters on position. For a boundary-value problem to be well posed for a parabolic equation (PE), unlike the wave equation, type, and even the number of boundary conditions (BCs) to be imposed at an interface, depends on the interface geometry and the PE itself. In this paper, the boundary conditions at staircase and sloping interfaces are considered for a class of PEs for which the operator square root is approximated by a rational function. Boundary conditions are *derived* from the PEs themselves thereby ensuring consistency between the equation and the BCs. In particular, energy-conserving PEs produce energy-conserving BCs. Physically, this procedure can be viewed as the substitution of an interface by an infinitesimally thin transition layer. The existence and uniqueness of solutions for the boundary-value problem for a wide-angle PE in the presence of interfaces are both illustrated by considering plane-wave reflection and transmission. Moreover, this fundamental problem provides insight into asymptotic accuracy of the parabolic approximation in media with piecewise-continuous parameters. [Work supported by NSERC.]

2:00

3pUW5. Simulated wave propagation in elastic environmental flat and sloping waveguides by a finite length ping time-dependent ray trace model. Elmer White (130 Moonraker Dr., Slidell, LA 70458)

A finite length ping ray trace program (FLPRT) has been extended which is now capable of simulating time-dependent wave motion in a complicated elastic bathymetric waveguide. From the original concept of sound waves being generated from an infinite set of point sources, FLPRT expands a finite set into the following ray theory including shear and compression logic in a half-space. Wave propagation is demonstrated in a constant velocity shallow-ocean environment for both flat and sloping bathymetries. Not only is the objective to compare the model-generated transmission loss calculations against those computed by EFEPE, an elastic parabolic equation model by Collins of ARL, but also show the time-generated path of the acoustic waves in both cases. Narrow angle-generated pings show nonmodal characteristics than those generated by wide-angle pulses which exhibit fluctuating waves predicted by normal mode theory. As in all cases, adiabaticity is assumed within each very

small incremental time step. A 3- to 5-min videotape will be shown on a VCR demonstrating wave motion in the presence of a flat and sloping wedge half-space exhibiting both shear and compressional waves on the half-spaces.

2:15

3pUW6. Flux-transfer amplitudes and the one-way coupled normal mode procedure. Michael F. Werby (Code 7181, NRL, Stennis Space Center, MS 39529) and Natalia A. Sidorovskaia (Univ. of New Orleans, New Orleans, LA 70148)

In a range-dependent environment, it is of interest to determine the way in which energy or flux proceeds in range. In the normal mode paradigm of propagation, it is possible to trace the energy flow according to specific modes. For a stratified environment there is no mechanism to alter the energy in each mode. It remains the same as long as the environment is range independent. However, for the range-dependent case, for which the coupled mode theory is adequate, it is possible to derive ''coupled mode amplitudes.'' The amplitude structure may then be studied in range to determine the manner in which energy traverses in the waveguide. This work illustrates what is meant by mode coupling amplitudes and the interpretative value in which the variation of these quantities in range adds insight in understanding a variety of mechanisms that alter or cause energy to proceed in a specific manner. [Work sponsored by ONR, the Naval Research Laboratory, and the University of New Orleans.]

WEDNESDAY AFTERNOON, 3 DECEMBER 1997 PRESIDIO ROOM, 3:00 TO 5:30 P.M.

Plenary Session, Business Meeting, and Awards Ceremony

Lawrence A. Crum, Chair
President, Acoustical Society of America

Business Meeting

Presentation of Certificates to New Fellows

Presentation of Awards

Science Writing Award in Acoustics for Journalists to Richard Wolkomir

Science Writing Award for Professionals in Acoustics to Michael J. Buckingham, Chad L. Epifano and John R. Potter

Distinguished Service Citation to Alice H. Suter

Honorary Fellowship to Robert W. Young

Silver Medal in Acoustical Oceanography to Herman Medwin

Silver Medal in Physical Acoustics to Robert E. Apfel

Silver Medal in Speech Communication to Patricia K. Kuhl

Wallace Clement Sabine Medal to Russell Johnson

Session 4aAAa

Architectural Acoustics: The Technical Committee on Architectural Acoustics Vern O. Knudsen Distinguished Lecture

Steven M. Brown, Chair
Steelcase, Inc., CD-5E-16, P.O. Box 1967, Grand Rapids, Michigan 49501

Chair's Introduction—9:00

Invited Paper

9:05

4aAAa1. Concert hall design: A consultant's perspective and retrospective. J. Christopher Jaffe (Jaffe Holden Scarbrough Acoust., Inc., 114A Washington Ave., Norwalk, CT 06854)

Architectural environments for symphonic performances can perhaps be described as one of the most complex problems in physical acoustics due to the multiplicity of sound source–path–receiver systems present in listening spaces. On the other hand, people performed and enjoyed listening to music in architectural and outdoor spaces long before the development of architectural acoustics as a scientific discipline and well before there were acoustical consultants. This paper describes one practitioner's historical perspective and analysis of the problem as it relates to public concert hall design and musical performance over the last several centuries and how the interrelationship between the two has actually fashioned the traditional symphonic environment. Based on this analysis, a methodology will be discussed to successfully replicate traditional symphonic environments utilizing psychoacoustic rather than geometric guidelines. The results of applying this methodology in a variety of building types over a period of 40 years, e.g., concert halls, recital halls, music pavillions, multiuse halls, etc., will be discussed. Both physical and electroacoustical solutions will be presented to support the author's original analytic premise.

10:00–10:30
Discussion

Session 4aAAb

Architectural Acoustics: Book Signing in Honor of ''The Sabines at Riverbank''

William J. Cavanaugh, Chair
Cavanaugh Tocci Associates, Inc., 327F Boston Post Road, Sudbury, Massachusetts 01776

Chair's Introduction—10:45

John Kopec will attend this session to sign copies of his book on the history of Riverbank Laboratories and the role of the Sabines (Wallace Clement, Paul Earls, and Hale Johnson) in the science of architectural acoustics.

Session 4aAB

Animal Bioacoustics: Biologically Inspired Acoustics Models and Systems II

Whitlow W. L. Au, Chair
Hawaii Institute of Marine Biology, P.O. Box 1106, Kailua, Hawaii 96734

Chair's Introduction—8:30

Invited Papers

8:35

4aAB1. Temporal and spectral information in echoes for biomimetic object recognition. Roman B. Kuc (Dept. Elec. Eng., Yale Univ., New Haven, CT 06520-8284)

Typical echolocation sounds produced by bats and dolphins have a wide bandwidth. This allows the echoes to contain information in the time and frequency domains. Temporal information includes the evolution of the energy in the perceived echo packet. However, if the initial echoes in a packet are weak, they may not exceed the perceptual threshold at a repeatable time, making temporal comparisons difficult. Spectral cues are more robust in that a time delay in an echo packet does not effect the power spectrum. However, because the emitted beam pattern and the reception patterns of the ears are frequency dependent, spectral comparisons are difficult. This paper describes methods for using temporal cues to aid in the spectral comparisons and methods for using spectral cues to aid in the temporal comparisons. The methods are illustrated using experimental results obtained with a biomimetic sonar mounted at the end of a robot arm that is used to recognize a variety of objects. [This research was funded through a grant from the NSF IRI-9504079.]

9:00

4aAB2. Real-time sonar classification of proud and buried targets using dolphinlike echolocation signals. Reid H. Shizumura (ORINCON Corp., 970 N. Kalaheo Ave., Ste. #C215, Kailua, HI 96734), Whitlow W. L. Au (Hawaii Inst. of Marine Biol., Kaneohe, HI), Gerald C. Moons (ORINCON Corp., Kailua, HI 96734), Paul E. Nachtigall (Hawaii Inst. of Marine Biol., Kaneohe, HI), Herbert L. Roitblat (Univ. of Hawaii, Honolulu, HI), and Robert C. Hicks (ORINCON Corp., Kailua, HI 96734)

A broadband active sonar system using dolphinlike echolocation signals and biologically inspired signal processing algorithms was developed to classify proud and buried targets in real time. A dolphin simulator transducer was attached to a bottom-crawling remotely operated vehicle and used to transmit a dolphin click (120-kHz center frequency, 39-kHz bandwidth, 50-μs duration) through seawater. Reflected target echoes were received and transmitted via cable to a shore-based computer system where the echoes were digitized at 1 MHz and subsequently processed. Two time-frequency representations of the echoes, one based on the short-time Fourier transform and the other on the Morlet wavelet, were processed in a hierarchical neural network system to derive target classifications. Echolocation returns were collected from six objects (cast iron pot, stainless steel sphere, glass jar, concrete tile, and coral rock) that were placed either on the ocean bottom or buried by bottom sediment. Echoes were separated into three target categories: (1) cast iron pot; (2) stainless steel sphere; and (3) a miscellaneous category consisting of the remaining four objects. Following supervised neural network training, the system was able to correctly identify 74%, 97%, and 88% of the category 1, 2, and 3 test echoes.

9:25

4aAB3. The adaptive silicon cochlea. Rahul Sarpeshkar (Bell Labs., Dept. of Biological Computation, Rm. 3L-404, 600 Mountain Ave., Murray Hill, NJ 07974, rahul@physics.bell-labs.com)

Low-power wide-dynamic-range systems are extremely hard to build. The biological cochlea is one of the most awesome examples of such a system: It can sense sounds over 12 orders of magnitude in intensity, with an estimated power dissipation of only a few tens of microwatts. An analog electronic cochlea that processes sounds over six orders of magnitude in intensity, while dissipating less than 0.5 mW, is described. This 117-stage, 100-Hz to 10-KHz cochlea has the widest dynamic range of any artificial cochlea built to date. This design, using frequency-selective gain adaptation in a low-noise, traveling-wave amplifier architecture, yields insight into why the human cochlea uses a traveling-wave mechanism to sense sounds, instead of using bandpass filters.

9:50–10:00 Break

10:00

4aAB4. Biologically inspired SCAT sonar receiver for 2-D imaging. James A. Simmons, Prestor A. Saillant, and Seth Boatright-Horowitz (Dept. of Neurosci., Brown Univ., Providence, RI 02912)

The SCAT (spectrogram correlation and transformation) model of biosonar in bats represents FM broadcasts and echoes as spectrographs with an integration time of 300–400 μs. The model then mimics signal processing by bats with two parallel computational paths: (1) determination of the delay of echo spectrograms using delay lines and coincidence detectors, and (2) deconvolution of the echo spectrum to determine delay separations smaller than the integration-time window [Saillant *et al.,* J. Acoust. Soc. Am. (1993)] to produce high-resolution echo-delay images. The new, binaural version of the SCAT model receives broadcasts and echoes at two ''ears'' and uses binaural algorithms to produce range/azimuth sonar images. The model incorporates several stages of binaural interactions that facilitate different functions including determination of azimuth itself, enhancement of target detection, and separation of adaptive imaging from these other functions. The model produces images that locate the principal glints in complex targets from only a few sonar broadcasts without having to sample many different target aspect angles. [Work supported by ONR and NSF.]

10:15

4aAB5. Target localization aboard a tone emitting robotic echolocation system with mobile receivers. V. Ashley Walker, Herbert Peremans, and John C. T. Hallam (Dept. of A.I., Univ. of Edinburgh, 5 Forrest Hill, Edinburgh EH1 2QL, Scotland, ashley@aifh.ed.ac.uk)

Using only *two* receivers, animals are able to determine the location of a target in *three* dimensions. Echolocating bats, which employ broadband calls, have access to spectral cues that may be used, in conjunction with interaural disparities of intensity (IIDs) and/or time (ITDs), to encode a 3-D target angle. This work addresses the question: what cues might a tone-emitting echolocator employ to determine target azimuth and elevation? To this end, a 6-degree-of-freedom robotic sensor head was built to investigate the direction cues that might be generated by the systematic pinnae scanning movements used by many species of tone emitting bats. The first strategy investigated was the use of receiver motion to rotate the SONAR horizon through discrete IID sampling positions so that a 3-D target angle can be resolved across IID readings taken at two or more receiver positions. Next, the change in amplitude measured by continuously scanning pinnae was used to create temporal cues, which vary systematically with the target angle. In this scheme, angular resolution depends upon scan speed (ms) rather than interaural separation (microsecond). Finally, the use of Doppler shifts to encode the target angle through the cosine law was investigated.

10:30

4aAB6. Estimating acoustic flow parameters for multiple echoes within a biologically motivated signal processing framework. Rolf Müller and Hans-Ulrich Schnitzler (Animal Physiol., Tübingen Univ., Morgenstelle 28, D-72076 Tübingen, Germany, rolf.mueller@uni-tuebingen.de)

It is hypothesized that proportional changes in echo amplitude and carrier frequency might aid obstacle avoidance in cf bats by means of acoustic flow perception. This approach appears to be feasible in principle for single targets, however, obstacle avoidance, unlike prey capture in midair, is invariably prone to constitute a multiple target problem, which has to be addressed in assessing the hypothesis' validity. In order to map a compound excitation pattern formed by superposition of responses to multiple echoes back into target space by means of an acoustic flow approach, instantaneous (carrier) frequency and amplitude together with their first-order derivatives must be recovered individually for each echo to be perceived as a separate entity. The primal sketch serving as a substrate for this operation is impoverished in that it is only preserving response envelopes. At the same time, the system design might capitalize on knowledge about the transfer functions of the filter bank's channels as well as at least a rough guess about the echo spectrum. Consequently, it is attempted to explain compound excitation patterns as a sum of superimposed responses to individual echoes, which are known for any given frequency location to a scaling constant.

10:45

4aAB7. The ecological approach toward ultrasonic perception. Herbert Peremans and John Hallam (Dept. of Artificial Intelligence, Univ. of Edinburgh, Forrest Hill 5, Edinburgh EH1 2Ql, UK)

Navigation seems to automatically entail the construction and use of environment models. In this paper it is shown how, in certain situations, the information provided by an echolocation system contains invariants that can be used directly to control the navigation behavior of an agent without the need for an explicit environment model. In particular, the case of an echolocating agent moving through an opening between two reflecting objects is analyzed. The transmit signal, duration 2 ms, is a frequency modulated sinusoid, hyperbolic frequency sweep from 100 down to 20 kHz. The processing performed on the raw measurement results, i.e., the signals picked up by the microphones, is modeled on the peripheral auditory processing of the mammal brain. The output of the model is an approximation of the short-time amplitude spectrum of the received signal. It will be shown that the control of the agent can be expressed solely in terms of invariant properties of the patterns appearing on the outputs of this model. This approach not only predicts the flight path of bats flying between vertical wires but it also significantly simplifies the control of a mobile robot moving between obstacles as will be shown.

11:00

4aAB8. Emergent collective behaviors in a simplified model of dolphin echolocation and predation. Thomas J. Hayward (Naval Res. Lab., Washington, DC 20375-5350)

A stochastic modeling framework for representing and simulating marine-mammal spatial distributions, collective motions, and vocalization occurrence times was previously developed [J. Acoust. Soc. Am. **101**, 3197(A) (1997)]. Emergent behaviors in these models include clustered and polarized motions simulating collective motions of marine mammals. The present work explores the potential of these models for representing and simulating more complex acoustic emission and collective motion behaviors related to short-range echolocation and predation by dolphins. A simplified predation scenario is constructed by including predator, prey, and nonprey species in the collective motion model. Echolocation is simulated by a simplified, sonar equation-based representation of the backscattered acoustic energy, with different backscattering characteristics for predator, prey, and nonprey individuals. Parameters determining echolocation signal types and collective motion characteristics are then allowed to vary in a random search that favors parameter values leading to higher prey capture rates. Emergent collective behaviors, including echolocation signals and prey pursuit and capture, are then observed in computer-generated video animations. The potential of the models for representating sound generated by groups of marine mammals and for interpreting observed marine mammal collective behavior is discussed. [Work supported by ONR Base funding at NRL.]

Session 4aBV

Bioresponse to Vibration/Biomedical Ultrasound and Physical Acoustics: Lithotripsy, Shock Waves and Bubbles and Tactile Estimation

Thomas J. Matula, Chair
Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105

Contributed Papers

8:30

4aBV1. *B*-scan ultrasound monitoring of cavitation acitivity in and around the kidney during shock wave lithotripsy. Robin O. Cleveland, Lawrence A. Crum (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105), David A. Lifshitz, Bret A. Connors, Lynn R. Willis, Andrew P. Evan (Indiana Univ. Med. School, Indianapolis, IN 46202), and James E. Lingeman (Methodist Hospital, Indianapolis, IN 46202)

Acoustic cavitation is thought to play a major role in both stone comminution and tissue injury during shock wave lithotripsy (SWL). Although ultrasound machines cannot directly observe cavitation events, remnant gas bubbles appear as echogenic regions. We treated young pigs in an unmodified Dornier HM3 and used a standard clinical ultrasound machine (Brüel and Kjær 3535) to image the kidney while SWL was being administered. We observed large regions of *in vivo* echogenicity within the renal pelvis (collecting system), which has not been reported before. When the ureter was blocked, up to 30% of the renal pelvis was covered with echogenicity. Clouds of echogenicity progressed through the collecting system with consecutive shots. The echogenic regions in the renal pelvis could be temporarily enhanced when nondegassed fluid was injected through a ureter catheter. Echogenicity was observed on the posterior and anterior surfaces of the kidney and appeared to occur at sites of subcapsular hematomas. The presence of echogenicity is consistent with the claim that cavitation plays an important role in both stone comminution and tissue damage. [Work supported by NIH and the Hunt Fellowship.]

8:45

4aBV2. Collapse of cavitation bubble induced during extracorporeal shock wave lithotripsy. Tetsuya Kodama and Kazuyoshi Takayama (Shock Wave Res. Ctr., Inst. of Fluid Sci., Tohoku Univ., 2-1-1 Katahira, Aoba Ward, 980-77 Japan, kodama@ifs.tohoku.ac.jp)

The interaction of air bubbles attached to gelatin surfaces, rat's livers, or rat's abdominal aortas with underwater shock waves was investigated to clarify the tissue damage mechanism by cavitation bubbles induced during extracorporeal shock wave lithotripsy. The overpressure of the shock wave was 10.2 ± 0.5 MPa ($n=4$). The initial bubble radii varied from 0.12–3.06 mm. The subsequent collapse of the bubbles was recorded by a high-speed framing camera. The liver cell damage was histochemically evaluated. The bubble attached to the gelatin or rat's liver surface migrates away from the surface with an oscillatory growth/collapse behavior after the shock wave interaction. The penetration depth of the liquid jet into the gelatin and the diameter of the subsequent damage pit on the surface depend on the initial bubble radius. The bubble near the rat's liver surface or aorta surface tends to show the same behavior as for the gelatin. The elongation and split of the nuclei in the liver parenchyma in the direction of the liquid jet and the increase in the cell density within the circumference of the injured region are histochemically revealed.

9:00

4aBV3. Damage of living tissue cells in water by shock waves and mathematical model. Masaaki Tamagawa (Faculty of Energy Sci., Kyoto Univ., Sakyo-ku, Kyoto 606-01, Japan) and Teruaki Akamatsu (Neyagawa, Osaka 572, Japan)

In this study, the damage experiments of the living tissue cells (red blood cells and cancer cells) [M. Tamagawa and T. Akamatsu, Proc. 20th Int. Shock Wave Symp., Pasadena, USA, 517–518 (1995)], and the experimental model for cells (microcapsules including dye stuff) by plane shock waves are executed. It is shown from the results that using a microcapsule is an effective method for sensing injury to living tissue cells. To explain these phenomena, the living tissue cells are modeled mathematically as one and two spherical elastic shells filled with liquid toward the shock wave. Using stationary and transient analysis of a spherical shell, the dynamic characteristics of the one living tissue cell and two with a mutual interaction are evaluated. The results show that damage to living tissue cells depends on: (1) the elastic modulus of the cell, (2) the bulk modulus of intracellular material, (3) the thickness of the cell membrane, (4) the distance between the cells on shock-induced damage, and (5) the rise time of the shock wave.

9:15

4aBV4. Overpressure reduces acceleration of thrombolysis due to ultrasound. E. Carr Everbach (Swarthmore College Eng. Dept., Swarthmore, PA 19081-1397, ceverba1@swarthmore.edu), Janice White, and Charles W. Francis (Univ. of Rochester, Rochester, NY 14642)

It has been shown that application of cw ultrasound at 1–8-W/cm2 SATA intensities and megahertz frequencies can accelerate the dissolution of blood clots when a clot-lysing agent is present. To determine whether or not a cavitational mechanism is responsible for ultrasound accelerated thrombolysis (UAT), a hyperbaric chamber was constructed that can apply 10 atm of static air overpressure to an *in vitro* exposure apparatus. Five hundred μl fibrin clots overlaid with a clot-lysing agent were exposed to 4-W/cm2 SATA ultrasound at 1 MHz and lysis quantified via release of bound radiolabel as a function of overpressure. More than half the acceleration due to ultrasound was removed, suggesting a bubble mechanism is responsible at least in part for UAT. Transmission and reflection of 20-MHz tone bursts of 0.5 μs duration provide evidence that bubble activity was reduced concomitantly during overpressure. [Work supported by NIH.]

9:30

4aBV5. Bubble collapse emissions suggest mechanism for transient response imaging. E. Carr Everbach (Swarthmore College, Eng. Dept., Swarthmore, PA 19081-1397, ceverba1@swarthmore.edu), Shouping Li, and Thomas R. Porter (Univ. of Nebraska Med. Ctr., Omaha, NE 68198-2265)

Transient response imaging (TRI) is the increase in apparent myocardial contrast with reduced image frame rate (PRF) when perfluorocarbon-filled echo-contrast agents are infused. TRI allows improved estimation of myocardial perfusion, but its mechanism is not known. Recent work sug-

gests that contrast microbubbles are destroyed by imaging at conventional frame rates (30–40 Hz), while less destruction occurs at cardiac-synchronized frame rates. Using a 20-MHz passive cavitation detector, acoustic emissions were recorded from within the myocardium of two anesthetized dogs *in vivo* during ultrasonic imaging with a clinical scanner. Strong correlations among root-mean-square detector output, videodensity, and PRF point to bubble emissions themselves as possibly responsible for the improved contrast. [Work supported by Nebraska AHA.]

9:45–10:00 Break

10:00

4aBV6. Nonlinear effects in HIFU propagation and attenuation in biological tissue. Vera A. Khokhlova, Oleg A. Sapozhnikov (Dept. of Acoust., Phys. Faculty, Moscow State Univ., Moscow 119899, Russia, vera@na.phys.msu.su), and Lawrence A. Crum (Univ. of Washington, Seattle, WA 98105)

For therapeutic applications of medical ultrasound, such as hyperthermia, acoustic surgery, and hemostasis, it is desirable, for precise targeting, to induce effective local absorption of the acoustic energy at a small focal area of the acoustic beam. A source with high-focusing gain can be used to achieve the most effective regime of focusing, in which nonlinear effects increase dramatically very close to the focal point. Sharp shocks are developed and the corresponding absorption rate increases at the focus without extra nonlinear attenuation in the prefocal region. In an earlier presentation [Khokhlova *et al.*, J. Acoust. Soc. Am. **98**, 2944 (1995)], a modified spectral approach was used that enables us to model propagation of nonlinear plane waves, including shocks, in biological tissue. Here, this method is extended for the description of high-intensity focused sound beams. The acoustic pressure field is calculated numerically for specific focusing gains and tissue absorption over a wide range of the source amplitudes. The effect of nonlinearity, focusing gain, absorption on the waveforms, amplification and spatial localization of the intensity, heating rate, and positive and negative peak pressure at the focus is discussed. [Work supported by ONR and CRDF.]

10:15

4aBV7. Remote generation of shear wave in soft tissue by pulsed radiation pressure. Victor G. Andreev, Vladimir N. Dmitriev, Oleg V. Rudenko (Phys. Dept., Moscow State Univ., Moscow, Russia 119899), and Armen P. Sarvazyan (Artann Labs., East Brunswick, NJ 08816, armen@mail.crisp.net)

Remote generation of shear waves in tissues by a focused ultrasonic beam is the basis of a new acoustic method of medical diagnostics, namely, shear wave elasticity imaging [A. P. Sarvazyan, U.S. Patent No. 5,606,971, 1997]. The feasibility of SWEI was demonstrated recently in experiments with optical and NMR detection of ultrasonically induced shear waves in tissue phantoms. In the present study an SWEI system with ultrasonic detection of shear waves was designed and tested. The shear wave was excited inside an inhomogeneous tissue phantom using radiation force generated by the reflection of ultrasound from internal inhomogeneity. A focusing transducer of 6 cm in diameter and with 6-cm focal length was used. The carrier frequency was 2 MHz, and the intensity was varied in range from 10–30 W/cm^2. The rectangular envelope was 0.2 ms in duration. Induced shear motion was evaluated using a 10-μs-long 3-MHz ultrasonic pulse by measuring the time delay for the pulse backscattered by moving inhomogeneities in the tissue. Maximum displacement in the propagating shear wave was over 10 μm, which agrees with theoretical predictions.

10:30

4aBV8. Nonlinear acoustic parameter of trabecular bone. Dimitri M. Donskoy (Davidson Lab., Stevens Inst. of Technol., Hoboken, NJ 07030, ddonskoy@stevens-tech.edu) and Alexander Sutin (Stevens Inst. of Technol., Hoboken, NJ 07030)

This work is an attempt to measure the nonlinear parameter of porous trabecular bone. Two different methods were applied. One method involves an AM ultrasonic signal transmitted through a bone submerged into water. A part of the signal energy transforms into the low-frequency (the frequency of modulation) signal due to the nonlinearity of the sample. The larger the nonlinear parameter of the sample, the more intensive such a transformation. This low-frequency signal is then picked up with a hydrophone and measured. The other method employs the effect of modulation of higher-frequency ultrasound by lower-frequency vibration applied directly to the bone. The same high frequencies and low frequencies were used in both tests, so the independently measured nonlinear parameters of the samples (cut from the same bone) could be compared. High dissipation of ultrasonic waves in trabecular bone in the frequency range above 100 kHz made it difficult to measure the nonlinear effects. The nonlinear parameter in the frequency range below 100 kHz was able to be measured reliably. Thus the nonlinear parameters were measured for three ultrasonic frequencies 26, 37, and 60 kHz modulated by the vibration frequencies 4.6 and 9.8 kHz. *In vitro* determined values of the nonlinear parameter for bovine bone used in these tests were in the range 80–120, which is an order of magnitude higher than for nonporous media (e.g., compact bone and liquid). [Work supported by NASA.]

10:45

4aBV9. Method of pathology recognition in bio-organs by multifrequency ultrasound signals. Lena I. Oboznenko (Acoust. and Acoustoelectron., Dept. of Electron., Ukrainian Natl. Tech. Univ., Kiev, Ukraine 252057)

Ultrasound methods of pathology recognition in bio-organs are proposed. They are based on the application of multichannel sounding signals (or wide band signals). The peculiarities of the proposed methods take into account the specific character of the biomedium in the ultrasound diagnostic. The nuances of the effective application in the ultrasound diagnostic of the following method of recognition are analyzed: application of the natural resonance frequencies of the bio-organs without signal form recollection (vessel, red corpuscles, leucocytes, etc.), by apportionment of complex poles of the Laplas transformation function (R. Prony's method), application of the stepped influence with organ boundary restoration by impulse reaction and by the algorithm of the form restoration with projection (J. Young's method and V. Bojarsky's method, J. Radon transformation), structure methods of the two-dimensional bio-organ form picture (the grammar method), multichannel irradiation of the bio-organ with its phase-frequency characteristic synthesis of three-dimensional bio-organ transformation by its Ess. in LF, MF, HF bands. The numerical data concerning the form compensation of the spherical and elliptical elastic objects and recognition of the red corpuscles and leucocytes in blood vessels are deduced. Frequency band: 1–10 MHz, feeler length: one period, discrete radiation: 1, 2, 4, 6, 8, 10 MHz.

11:00

4aBV10. Tactile estimation of roughness and size. Ronald T. Verrillo, Stanley J. Bolanowski (Inst. for Sensory Res., Syracuse Univ., Syracuse, NY 13244), and Francis P. McGlone (Port Sunlight Lab., Bebington Wirral L63 3J2, England)

Subjects estimated the subjective magnitude of roughness and the size of steel balls by the method of absolute magnitude estimation. Using the index finger and thumb, estimates were made of 11 grades of sandpaper with particle diameters ranging from 16–905 μm. The results describe a

power function and show spatial summation when stimuli are applied simultaneously to two sites (finger and thumb). The same subjects judged the size of nine steel balls ranging in diameter from 0.4–6.4 mm. Estimates were made with combinations of the index finger, volar forearm, and thumb. Again, all of the results were power functions and spatial summation of tactile stimulation was evident. The results are discussed in terms of theoretical considerations and practical application. [Work supported by Unilever and NIH, PO1DC00380].

THURSDAY MORNING, 4 DECEMBER 1997 FORUM ROOM, 8:55 TO 11:35 A.M.

Session 4aEA

Engineering Acoustics: Acoustic Systems Designed for Harsh Environments

Fred C. DeMetz, Chair
Allied Signal Ocean Systems, 15825 Roxford Street, Sylmar, California 91342

Chair's Introduction—8:55

Invited Papers

9:00

4aEA1. Reliable high-temperature electronics for instrumentation and data acquisition. Ben Gingerich (Honeywell Solid State Electron. Ctr., 12001 State Hwy. 55, Plymouth, MN 55441)

Honeywell is developing and has introduced the first products of the HTMOS™ product line, which is targeted for use in systems operating in severe high-temperature environments. The HTMOS™ line of electronic products incorporates Honeywell's oxide-isolated high-temperature process and high-temperature circuit design methodologies. These components are being developed to meet applications in instrumentation, data acquisition, and control where hostile environments require a large temperature range of −55 to +225 °C with an extended operating life of 5 years at elevated temperatures. Honeywell has introduced a quad operational amplifier, quad analog switch, an 8-bit microcontroller, and a 32 K×8 SRAM. Additional products are planned for introduction in 1997. This paper will discuss the high-temperature features of the silicon on insulator (SOI) process technology and product design approaches for reliable high-temperature electronic components. The paper will also discuss the existing product performance in a high-temperature environment.

9:25

4aEA2. Pressure-tolerant network interface cards for sensor arrays. John Walrod (Planning Systems, Inc., 21294 Johnson Rd., Long Beach, MS 39560, john_walrod@psilongbeach.com)

Pressure-tolerant ATM and Ethernet network interface cards (NICs) were developed for towed array data acquisition and telemetry. The card functions include sensor signal conditioning, analog-to-digital conversion, network protocol processors, and network interface transceivers. Design and fabrication techniques were developed to achieve space constraints, power constraints, and operation at 2500-psi pressure in oil or saltwater submersion without the use of pressure vessel housings. The development included the testing of candidate electronic components and NIC assemblies for pressure, temperature, and oil tolerance. Components and assemblies which failed testing were analyzed, and design rules were established for future designs. Generally, the elimination or minimization of voids by proper component selection and fabrication was found to be critical. Pressure-tolerant fabrication techniques were developed such as temperature cycle screening, encapsulation in low-viscosity polyurethane, pressurized curing, and compression sealed wires. High-density printed circuit boards and minimalist circuit designs were used to minimize NIC size. Low-voltage components, selective component location, and thermally conductive packaging were used to minimize operating temperatures. Commercially available components and open standard interfaces were used to minimize costs.

9:50

4aEA3. Comparison of a thermally biased PMN-PT transducer and a conventional PZT-based underwater source. Scott A. Hudson (AlliedSignal Ocean Systems, 15825 Roxford St., Sylmar, CA 91342)

Recent research into PMN (lead magnesium niobate)-based electrostrictive materials have led to compositions with strain rates significantly higher than that of piezoelectric ceramics. The large strain rate makes it ideal for use in high-power SONAR transducers. However, unlike piezoelectric ceramics, the properties of PMN vary nonlinearly with temperature. The highest performance PMN compositions exhibit optimal characteristics at temperatures near 85 °C. AlliedSignal has solved the engineering challenges of implementing a high-temperature PMN composition in a SONAR transducer and demonstrated the performance at NUWC Lake Seneca test facility. This paper discusses the properties of high-temperature PMN materials and presents the test results of a PMN transducer compared with a similar transducer which utilizes conventional piezoelectric ceramic. The PMN transducer produced roughly 6 dB higher source level than the PZT transducer. Electromechanical coupling and bandwidth of the two transducers was similar.

10:15–10:30 Break

10:30

4aEA4. Development of an advanced vibratory source for borehole seismology. Bjorn N. P. Paulsson (Paulsson Geophysical Services, Inc., 1300 Beach Blvd., La Habra, CA 90631-6374)

Single-well seismology, reverse vertical seismic profiles (RVSP's) and Crosswell seismology are three new seismic techniques for obtaining much higher resolution images of oil and gas reservoirs than what is currently obtainable with surface seismic techniques. Borehole seismology involves inserting both the source and the seismic receivers in oil or gas wells. In the past these methods have been limited to short distances between source and receivers in shallow wells. In 1997 a new borehole seismic source and receiver system has been deployed with a capability of handling source–receiver spacings of more than 2000 m in rock with a Q value of more than 20, and well depths as great as 6500 m. This advanced borehole seismic data acquisition system consists of a powerful, clamped, swept-frequency, vibratory source which is nondestructive, a multilevel receiver string of clamped, three-component geophones, and an acquisition system for reverse VSP's. The downhole vibrator produces high-quality S waves, with controlled polarization, as well as P-wave direct arrivals and reflections. This paper discusses the design and performance of this system.

10:55

4aEA5. Acoustic systems for well logging at extreme temperature and pressure. W. D. Squire and H. J. Whitehouse (Linear Measurements, Inc., 4174 Sorrento Valley Blvd., San Diego, CA 92121)

Various acoustic systems will be described briefly that have been studied and developed by one or both of the authors for use in well logging. Most of these systems have been developed with the goal of operating at extreme temperature (in some cases as high as 200 °C). A few of them have been developed with the additional goal of operating at extreme pressure (in some cases as high as 20 000 psi). A list of the systems to be discussed follows: a wireless telemetry system to transmit data from the bottom of the well to the surface, utilizing the drill pipe as an acoustic wave guide with zero-order torsional waves on the drill pipe as the data carrier; an oil and gas well passive listening device or an acoustisonde (a stethoscope or, as called by the petroleum industry, a noise tool); a wireless telemetry system to transmit data from the bottom of the well to the surface, utilizing the drill pipe as an acoustic wave guide with acoustic pulses in the drilling fluid (drilling mud) inside the drill pipe as the data carrier, with a data rate greater than existing mud pulse systems; an acoustic borehole televiewer technology with an order of magnitude finer resolution than existing systems; and advanced pressure release materials for use in acoustic transducers at extreme temperature and pressure.

Contributed Paper

11:20

4aEA6. Perspective directions of vector-phase methods application in acoustic-seismic environmental monitoring. Valery B. Mit'ko (Dept. of Marine Information Systems, Electrotechnical Univ. 5, Prof. Popov St., St. Petersburg, 197376, Russia)

Estimation of harmful influence parameters of acoustic-seismic fields is important in environmental complex monitoring to decide such problems as the determination of infrasound harbinger of tsunami, earthquake. Such enormous engineering structures as dams, power stations heat changers, and main pipelines are also infrasound sources and should be under control. Theoretical methods of sound field estimation are noneffective in low and underlow frequencies because of the simplifying assumptions. The mathematical model of vector-phase methods is based, for example, on solving for the integral on the closed surface. It permits one to define vector-kinematic characteristics of an acoustic field and its spatially-temporary distribution in controllable volume. The development of elastic wave-field parameter measurements and the creation of multicomponent receiver parks allow one to carry out background parameter measurement or responses to pulse influences with the purpose of heterogeneity localization in sedimental or radical layers of the researched area with minimal expense of energy and means for nature measurement realization. The presence of spectral composites as a result of local heterogeneities in the research area was established. It was noted that stream of power measurements provide the increasing signal/noise ratio up to 16–20 dB in different conditions.

THURSDAY MORNING, 4 DECEMBER 1997 COUNCIL ROOM, 10:30 A.M. TO 12:00 NOON

Session 4aED

Education in Acoustics: Take Fives—Sharing Ideas for Teaching Acoustics

Thomas D. Rossing, Cochair
Physics Department, Northern Illinois University, DeKalb, Illinois 60115

Uwe J. Hansen, Cochair
Department of Physics, Indiana State University, Terre Haute, Indiana 47809

Do you hava 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 4aNS

Noise and Engineering Acoustics: Noise Modeling and Outdoor Sound Propagation

Victor W. Sparrow, Chair
Graduate Program in Acoustics, Pennsylvania State University, 157 Hammond Building, State College, Pennsylvania 16804

Contributed Papers

8:30

4aNS1. The application of geographic spatial analysis in modeling traffic noise propagation away from a freeway. John C. Bennett (Dept. of Geography, San Diego State Univ., 5500 Campanile Dr., San Digeo, CA 92182-4493)

This paper demonstrates the use of a geographic information system (GIS) to perform a spatial analysis of traffic noise away from a freeway. The more traditional analysis procedures used by the Federal Highway Administration (FHWA, 1978 and 1982) employ a noise model that is range limited and not designed to handle large-scale topographic variation near a freeway noise source. A digital elevation map of a portion of a freeway in San Diego County that includes a mesa and valley landscape is incorporated into a GIS where a spatial analysis of ray path geometry, substrate, and vegetation is used to evaluate measured noise levels of traffic. Preliminary results indicate that a GIS model of traffic noise provides a more realistic basis for predicting noise impacts, particularly at longer ranges in areas of variable topography.

8:45

4aNS2. Six-year study of IMSA road racing noise. Gordon Bricken and Christopher Jean (Gordon Bricken and Assoc., Inc., 1621 E. 17th St., Santa Ana, CA 92705-8518)

The Del Mar Fairgrounds, San Diego, hosted the IMSA Grand Prix for 6 years from 1987–1993. The project provides a unique long-term study of professional road racing noise in a unique setting, a narrow valley just east of the coastline and within an environmentally sensitive lagoon. Sponsors were required to perform community noise monitoring at selected locations in the adjacent hills and lagoon. Both human and bird responses were examined. The monitoring documented the hour by hour variations in the maximum and average noise levels at up to 12 separate locations. The results indicated the thresholds which would trigger community response, the sensitivity of the response to diurnal and short-term wind shift, the effect of the inversion layer, the effects on bird patterns, and the statistical distribution of vehicle noise levels for the IMSA class vehicle. Basic findings revealed community response was triggered when levels exceeded 70 dBA Lavg. Downwind levels tended to increase at 1 dBA/mph over 5 mph. When the track was within an inversion, levels increased 10 dBA. The mean vehicle level at 15 m was 105 dBA with a 4-dBA standard deviation.

9:00

4aNS3. Nine-year study of NHRA drag racing noise. Gordon Bricken and Christopher Jean (Gordon Bricken and Assoc., Inc., 1621 E. 17th St., Santa Ana, CA 92705-85181)

Pomona International Raceway, Pomona, California is the premier venue for drag racing sanctioned by the National Hot Rod Association. Community noise was studied for 9 years from 1984–1993. The project provided a unique long-term study of professional drag racing noise at 14 locations in a built-up community adjacent to the starting line. The object was to establish a basis for documenting the effects of noise mitigation's planned for a renovated facility. The study revealed startling diurnal variations in noise, year to year variations, and the wide variety of emission characteristics of racing types. The simultaneous community and trackside locations permitted accurate correlation of events and established the basis for predictive models. The large data base permitted creation of a benchmark location by which the construction mitigation measures could be verified in a statistically valid manner. The study also revealed that the data spread is so wide that only multiple event sampling will provide an accurate picture of the community noise levels.

9:15

4aNS4. Effectiveness of random edge barriers: Further studies. Eric J. Rosenberg and Ilene J. Busch-Vishniac (Dept. of Mech. Eng., Univ. of Texas, Austin, TX 78712, erosenberg@mail.utexas.edu)

Previously, the idea of introducing a random height fluctuation to the conventional straight edge of a noise barrier was discussed as a way to increase insertion loss and improve acoustic performance. Here, an exhaustive set of experiments is discussed to more accurately discern the effects of the random edge. Experimental results were obtained using a spark source and a plywood barrier with interchangeable metal edges serving as the various edge profiles. Unlike the past experiments, which reported peak-to-peak sound pressures, results were obtained for the energy spectral density. The results indicate that the insertion loss data obtained with a random edge are not solely a function of Fresnel number as they are with a straight edge, and that local effects may dominate when considering the energy measured at a point in space. [This work was supported by a Texas Advanced Technology grant.]

9:30

4aNS5. Approximation to study the acoustic field diffracted in the shadow of a half-infinite barrier by the boundary element method. Antonio Sanchis, Albert Marín, José Romero, and Alicia Giménez (Appl. Phys., Dept., High Tech. School of Industrial Eng., Polytech. Univ., P.O. Box 22012, 46080 Valencia, Spain)

The application of the BEM (direct form) for the external domain, with a lineal combination of Helmholtz's equation integral and their derivative (Burton, Miller) makes possible the calculation of the unique solution for each number of waves. The fundamental characteristics of this model that works with lineal sound source are: (1) The calculation surface is the topographical profile adapted with adjacent plans, discertizated in elements defined by two extreme nodes and one central node. (2) The acoustic impedance of each element is determined starting from the Delany equation. (3) The unknowns are the sound field in each node ϕ_i. (4) A form function ϕ is determined starting from the ϕ_i of each node. (5) Helmholtz's operators are obtained by integrating the ϕ function in each element. (6) By solving the equation system with these operators, the values of the ϕ_i, and the function ϕ, are obtained. (7) By integrating Helmholtz's equation integral starting from ϕ, the value of the sound field in each point of the space is obtained. Finally, the model is applied to several simple barriers for which there are experimental values. The comparison of the obtained outputs shows the validity of the calculation.

9:45–10:00 Break

10:00

4aNS6. The calculation of curved path sound propagation in temperature inversion conditions. Robert L. Bronsdon (Walt Disney Imagineering, 1401 Flower St., Glendale, CA 91221)

Sound propagation models have typically been used to develop noise control solutions for a wide variety of troublesome sources from highways, power plants, and airports to stadiums and outdoor amphitheaters. Almost all of these models assume that there is no variability in the atmospheric properties of the propagating medium but this assumption is almost invariably false. Predictions from these models are usually validated during the day when propagation conditions are favorable, producing lower levels than the models calculate. The problems usually come at night, when temperature and wind variations in the atmosphere create propagation conditions which are not favorable, resulting in much higher noise levels in the community, often on the order of 15–20 dB. This paper describes work which has been undertaken by the Walt Disney Company in an effort to produce a more precise computer model for these unfavorable conditions. The model developed is based on Gaussian beams, a hybrid technique developed to retain some of the fundamental aspects of ray tracing while incorporating some of the more advanced GFPE concepts. The fundamental code, developed by Ken Gilbert and Xiao Di, is being integrated into SOUNDPLAN, a commercially available sound propagation computer program.

10:15

4aNS7. Wide-angle parabolic equation in moving inhomogeneous media. V. E. Ostashev (Dept. of Phys., Box 30001, New Mexico State Univ., Las Cruces, NM 88003-8001, vostashe@nmsu.edu), Philippe Blanc-Benon, and Daniel Juvé (Ecole Centrale de Lyon, 69131 Ecully, France)

The wide-angle parabolic equation has been used widely to predict a sound field in an ocean with variations in the adiabatic sound speed c. The same equations with c replaced by the effective sound speed $c_{\text{eff}}=c+v_x$ has recently been adopted for calculations of sound propagation in a stratified and/or turbulent atmosphere. Here, v_x is the wind velocity component in the direction from a source to a receiver. However, as is shown in the presentation, such a replacement does not allow one to correctly take into account the effects of the medium velocity **v** on sound propagation and scattering in a wide-angle approximation. Furthermore, the effects of fluctuations in the density ρ on sound scattering are also omitted in this approach. In the presentation, a correct wide-angle parabolic equation in media with variations in c, ρ and **v** and its Padé (1,1) approximation are derived. The difference in numerical predictions of sound propagation in a stratified and turbulent atmosphere, based on the equation derived and that used previously, is demonstrated. [This material is based upon work supported in part by the U.S. Army Research Office under Contract No. DAAH04-95-1-0593.]

10:30

4aNS8. Sound propagation through a turbulent jet: Experiment and theory. V. E. Ostashev (Dept. of Phys., Box 30001, New Mexico State Univ., Las Cruces, NM 88003-8001, vostashe@nmsu.edu), Philippe Blanc-Benon, and Daniel Juvé (Ecole Centrale de Lyon, 69131 Ecully, France)

The coherence functions of a spherical sound wave after passing through a turbulent jet were measured for different values of sound frequencies and mean speed of the jet [Ph. Blanc-Benon, Ext. Rev. Cethedec-Ondes Signal, No. 79-1984 (1984)]. In the presentation, the experimental data obtained are compared with the theoretical predictions based on the previous theories of sound propagation in turbulent media with medium velocity fluctuations [e.g., E. H. Brown and F. F. Hall, Rev. Geophys. Space Phys. **16** (1), 47–110 (1978)] and with the predictions of the theory recently developed [V. E. Ostashev, Waves Random Media **4**, 403–428 (1994)]. The differences between these theories and their results for the coherence function of a spherical wave are explained. It is also shown that the predictions based on the previous theories do not agree with the experimental data. On the other hand, the predictions of the theory developed by V. Ostashev fit the measured coherence functions fairly well. [This material is based upon work supported in part by the U.S. Army Research Office under Contract No. DAAH04-95-1-0593.]

10:45

4aNS9. Focusing of sonic boom noise penetration into a homogeneous wavy ocean: Complex surfaces and wavelength comparisons. Judith L. Rochat and Victor W. Sparrow (Penn State Univ., P.O. Box 30, State College, PA 16804, rochat@sabine.acs.psu.edu)

The last decade has seen a revived interest in the study of sonic booms; this is due to an upcoming new breed of supersonic passenger aircraft along with a heightened awareness of marine environmental issues. Sonic boom noise penetrating into a *flat, homogeneous* ocean is a topic several researchers have already addressed. The primary goal of the authors' research is to study the effects of *realistic* ocean features on noise penetration. These features include a wavy ocean surface and bubbles near the surface. This presentation will focus on a wavy ocean surface, specifically on somewhat complex surface profiles and on sonic boom effective wavelength/ocean wavelength comparisons. Results using finite difference simulations indicate that a somewhat complex wavy ocean surface profile slightly augments the underwater pressure field (as compared to a flat surface); this result was also found when studying simple surface profiles. Some of the trends observed, however, were not as obvious as seen when studying the simple surface profile. This prompted a study on how the sonic boom effective wavelength compared to the ocean wavelength affects the amount of augmentation to the underwater pressure field. [Work supported by the NASA Langley Research Center.]

11:00

4aNS10. Methods of correlating measured sound levels from an aircraft noise monitoring system with aircraft operations. Steve Alverson (Harris, Miller, and Hanson, 945 University Ave, Ste. 201, Sacramento, CA 95825)

The Port of San Diego, owner and operator of San Diego International Airport (SAN), maintains a comprehensive aircraft noise control program that is designed to minimize noise impacts from aircraft operations at San Diego International Airport—Lindbergh Field. One element of this program involves continuous monitoring of noise at 24 remote monitoring terminals in the neighborhoods surrounding SAN and the acquisition of aircraft flight track data. This paper explores methods used by the SAN Noise Information Department staff to correlate, using algorithms in the aircraft noise and operations monitoring system (ANOMS), measured aircraft sound levels with actual aircraft operations. The paper will discuss the use of sound level rise and decay times, frequency filters, timing of noise events between monitors, passive secondary surveillance radar (PASSUR) to acquire aircraft flight tracks, and aircraft situation display (ASD) information to identify aircraft only noise events. The paper will also discuss how measured aircraft noise levels are compared to computed aircraft noise levels to verify the accuracy of the annual community noise equivalent level (CNEL) contours. The CNEL contours are submitted to the State Department of Transportation in response to the State of California aircraft noise regulation reporting requirements.

Session 4aPP

Psychological and Physiological Acoustics: Temporally Dynamic Psychophysics and Physiology

Sid P. Bacon, Chair
Department of Speech and Hearing Science, Arizona State University, Box 871908, Tempe, Arizona 85287-1908

Contributed Papers

8:45

4aPP1. Rhythmic masking release: A paradigm to investigate auditory grouping resulting from the integration of time-varying intensity levels across frequency and across ears. Martine Turgeon and Albert S. Bregman (Dept. of Psycho., McGill Univ., 1205 Dr. Penfield Ave., Montreal, QC H3A 1B1, Canada)

The perceptual organization of sequences of short-duration 200-Hz-wide modulated noise bands was investigated. Sequences with one of two possible rhythms were presented over headphones or in a semi-circular array of 13 speakers. The ''rhythm'' was embedded in an irregular sequence of ''maskers'' with the same frequencies and energy. Adding ''flankers'' in other critical bands, synchronous with the irregular maskers released the rhythm from masking. RMR as measured by the correct identification of the rhythm and its perceived clarity was assumed to be contingent upon the perceptual grouping of the maskers and flankers. Using a 2-AFC procedure with a four-point clarity rating for each alternative, the effect of the following relations between the maskers and flankers on their simultaneous grouping was investigated: stimulus onset asynchrony (SOA), frequency separation (ΔF), commonality of microstructure, and angular separation of their sources ($\Delta\theta$). RMR decreased significantly with SOA, $\Delta\theta$, differences in microstructure and with ΔF for spatially separated maskers and flankers. RMR will be compared to other paradigms which have been claimed to result from a cross-spectral binaural analysis of the temporal envelope, such as dichotic CMR and/or of the fine temporal structure such as BMLD [Hall *et al.*, J. Acoust. Soc. Am. **83**, 1839–1845 (1988)].

9:00

4aPP2. Effect of temporal gaps on informational and energy-based masking. Toktam Sadralodabai, Donna L. Neff, and Traci R. Gleason (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131)

The effect of a temporal gap on detecting a 1000-Hz tone in the presence of 60-dB SPL simultaneous maskers was examined. Ten-component, random-frequency maskers and broadband-noise maskers were used in a 2-AFC adaptive task. Random-frequency components were drawn from 300 to 3000 Hz, excluding a 160-Hz band around the signal. Temporal gaps of 10, 20, 40, 80, and 160 ms were tested, positioned either at the onset, center, or offset of either the signal or the masker. Without gaps, both signal and masker durations were 200 ms. To maintain equal energy across all conditions, level compensation was applied when gaps were employed. For temporal gaps in either the multicomponent or noise masker, masked thresholds consistently decreased as gap duration increased. Gaps in the masker appeared to provide a temporal window for detection of the signal. However, for gaps in the signal, masked thresholds decreased with the multicomponent masker, but remained constant with broadband noise masker. With multicomponent maskers, the gaps appeared to reduce informational masking by perceptually segregating the signal from the masker. With broadband noise maskers, there was little informational masking and therefore the temporal gaps did not improve performance. [Work supported by NIDCD.]

9:15

4aPP3. Reducing the effects of masker uncertainty with harmonicity and onset/offset synchrony. Eunmi L. Oh and Robert A. Lutfi (Dept. of Psych. and Dept. of Commun. Disord., Univ. of Wisconsin, Madison, WI 53706)

Detection thresholds for a tone in an unfamiliar tonal pattern can be greatly elevated under conditions of masker uncertainty [E. Oh and R. A. Lutfi, J. Acoust. Soc. Am. **101**, 3148 (1997)]. Two experiments were conducted to determine whether harmonicity and onset/offset synchrony of masker tones can reduce the detrimental effect of masker uncertainty. Masker uncertainty was introduced by randomly varying the frequencies and amplitudes of masker tones on each presentation. In experiment 1, maskers were composed of 2–49 tones that were multiples of 200 Hz, not including 1000 Hz. Anywhere from 2–10 dB less masking was observed for an inharmonically related signal at 1047 Hz than for a harmonic signal at 1000 Hz. In experiment 2, the onsets and offsets of masker tones covaried or varied independently of each other. As much as 30 dB less masking was obtained for the 1000-Hz tone signal when masker tones covaried in time. The results support the idea that harmonicity and onset/offset synchrony cause the signal to be perceptually segregated from the masker, thereby resulting in less masking.

9:30

4aPP4. Spectral integration and the detection of tones in modulated and unmodulated noise. Sid. P. Bacon, Nicolas Grimault, and Jungmee Lee (Psychoacoust. Lab., Dept. of Speech and Hearing Sci., Arizona State Univ., Tempe, AZ 85287-1908, spb@asu.edu)

The intent of the present study was to determine whether spectral integration is similar in modulated and unmodulated noise. The signal consisted of one sinusoid of a triplet, or all three sinusoids together. Four sets of triplets were used. The masker consisted of one or three 100-Hz-wide bands of noise centered at the sinusoidal component(s) of the signal. They were unmodulated or sinusoidally amplitude modulated (in phase) at a rate of 8 Hz and a depth of 1.0. Thresholds were measured adaptively for each sinusoid in a triplet, and then for all three together, with the relative level of each determined by the threshold differences when measured individually. In unmodulated noise, the improvement in threshold (from one to three sinusoids) was about 2 dB, consistent with a $\sqrt{3}$ improvement. In modulated noise, however, the improvement was often greater than that (5 dB). This greater improvement could not be explained by shallower psychometric functions in the modulated condition, as the psychometric functions for the detection of the individual sinusoids were similar for both backgrounds. Instead, it is likely due to the comodulated nature of the three masker bands. [Work supported by NIDCD.]

9:45

4aPP5. Auditory temporal microstructure: Evidence of under- and over-shoot at onset and offset of a narrow-band noise masker measured in an off-frequency masked detection task. Craig Formby, Sarah H. Ferguson[a] (Div. of Otolaryngol.—HNS, Univ. of Maryland School of Medicine, 16 S. Eutaw St., Ste. 500, Baltimore, MD 21201, cformby@surgery2.ab.umd.edu), and Michael G. Heinz (Harvard-MIT, Cambridge, MA 02139)

Detection thresholds were measured as a function of the temporal position of a 5-ms sinusoidal signal presented within one of two 500-ms observation intervals. Both intervals contained a 400-Hz-wide noise masker, logarithmically centered at 2500 Hz, that was gated on at 150 ms and off at 350 ms within each interval. The signal and masker were both gated with 2-ms linear ramps. The level of the signal was tracked adaptively in blocks of forty 2I 2AFC trials to estimate a 70.7% correct detection threshold. Thresholds were measured at 43 temporal signal positions within the observation interval, with detailed sampling around the onset and offset of the masker to assess temporal edge effects. Experiments to date with four listeners have focused on signals presented at 1900 or 2100 Hz. The temporal microstructure of the detection functions has characteristically revealed a complex pattern of under- and over-shoot (i.e., enhanced and diminished detection, respectively). These effects, which have been observed at both the onset and offset of the masker for virtually all masker spectrum levels between $N_0 = 0$ and 70 dB, may accentuate the temporal edges of the masker and, in general, enhance detection of acoustic onsets and offsets. [a]Currently at Dept. of Speech and Hearing Sci., Indiana Univ.

10:00

4aPP6. Effects of fine structure on onset behavior of distortion products in humans. Carrick L. Talmadge (Dept. of Phys., Purdue Univ., West Lafayette, IN 47907), Glenis R. Long, and Arnold Tubis (Purdue Univ., West Lafayette, IN 47907)

There is accumulating evidence that the fine structure of distortion product otoacoustic emissions (DPOAEs) in humans originates from an interference between two cochlear sources, one at the ''overlap'' site of the two primaries, and the other at the best frequency for the DPOAE. It is shown that the DPOAE onset behavior measured in humans, after the turn on of an external tone, is consistent with previous evidence for a ''two source'' origin of DPOAE fine structure [reviewed in Mills, J. Acoust. Soc. Am. **102**, 413–429 (1997)]. A variation of Whitehead's methodology [J. Acoust. Soc. Am. **100**, 1663–1679 (1996)] was used to measure the initial DPOAE buildup. The $f1$ primary was on continuously for a given measurement and the $f2$ primary was pulsed (square wave, 50% duty cycle) at a rate of 5 Hz. The DPOAE had an initial buildup, followed either by further growth of the DPOAE, when the DPOAE frequency was near a fine-structure maximum, or by decay of the DPOAE, when the DPOAE frequency was near a fine-structure minimum. This pattern of secondary growth/decay is shown to be further evidence favoring the two-source model of DPOAE fine structure.

10:15–10:30 Break

10:30

4aPP7. The ''center-of-gravity'' effect in dynamic tones. Jayanth N. Anantharaman, Ashok K. Krishnamurthy (Dept. of Elec. Eng., Ohio State Univ., Columbus, OH 43210, jayanth+@osu.edu), and Lawrence L. Feth (Ohio State Univ., Columbus, OH 43210)

Earlier work [e.g., Xu *et al.*, J. Acoust. Soc. Am. **101**, 3149(A) (1997); Chistovich, *ibid.* **77**, 798–805 (1985)] has demonstrated the spectral center-of-gravity (CoG) effect. The studies have largely concentrated on steady-state signals. Recently, Lublinskaya (1996) has investigated the CoG effect in dynamic synthetic speech. In the study presented here, the CoG effect in dynamic two-tone complexes was investigated. The frequencies of the two tones were held constant while their amplitudes were varied linearly: the amplitude of the lower frequency tone was decreased while that of the higher frequency tone was increased. This causes a variation in the spectral CoG, although the locations of the frequency content remain constant with time. Subjects matched the slope of a linear glide (target) having the same center frequency as the two-tone complex (standard) using a double-staircase procedure. The experiment was carried out for various center frequencies (500–2000 Hz) and frequency separations (0.5–8 ERBs). The center frequency was also roved between trials. Matching experiments were also carried out with another linear glide as the standard. The results of these experiments will be presented and the center-of-gravity effect in these dynamic two-tone complexes will be discussed. [Work supported by AFOSR and OSU-GSARA.]

10:45

4aPP8. Discriminability of correlated bursts of reproducible noise. Susan M. Fallon (Dept. of Psych., Univ. of Alaska, 533 East Pioneer Ave., Homer, AK 99603)

Discriminability of pairs of 100-ms bursts of noise was measured using a same–different psychophysical method. In the first condition, the correlation between bursts in a pair was fixed at 1.0 during each same trial. On each different trial, bursts were identical except for τ ms of independent noise located at either the beginning, the middle, or the end of the pairs. The correlation between bursts presented on each different trial was increased from 0.0 by decreasing the duration of τ. The second condition was the converse of the first; the correlation between bursts presented on each different trial was fixed at 0.0 and the correlation between bursts presented on each same trial was decreased from 1.0. As the τ ms of either independent (condition 1) or the perfectly correlated (condition 2) noise is moved from the beginning to the end of bursts, discriminability increases in the first condition and decreases in the second condition. These results suggest that a ''difference'' is a more salient cue in the discrimination task than a ''sameness'' and are discussed in terms of a correlational model. [Work supported by NIH and AFOSR.]

11:00

4aPP9. Sequential interactions in the discrimination of frequency increments co-occurring with irrelevant increments in frequency, duration, or level. Blas Espinoza-Varas and Hyunsook Jang (Commun. Sci. and Disord., Univ. of Oklahoma Health Sci. Ctr., Oklahoma City, OK 73190, blas-espinoza-varas@uokhsc.edu)

Discrimination of target frequency increments co-occurring with irrelevant increments in frequency, duration, or level was studied. *Pairs* (t_1,t_2) of 1500-Hz, 80-ms pure tones separated by 20–120 ms of silence (ITI) were presented in a three-interval, two-alternative, forced-choice (3I/2AFC) task. On each trial, a ''standard'' *pair* of tones (interval 1) was followed by two ''comparison'' *pairs* (intervals 2 and 3), with 500 ms of silence between observation intervals. One (randomly chosen) comparison *pair* contained a target increment in the frequency of either t_1 *or* t_2; the size of the target increment was controlled by adaptive rules. In addition to the target increment, both comparison stimuli contained identical irrelevant increments in either the duration, frequency, or level of either t_1 or t_2; there was no correlation between the observation intervals containing target and irrelevant increments. Listeners had to determine which observation interval contained the target increment. Discrimination of t_1 frequency was degraded by irrelevant increments in t_2 frequency or t_2 level, particularly at the shortest ITI of 20 ms; prolonged discrimination training lessened the effects. Discrimination of t_2 frequency was not affected by irrelevant increments. [Work funded by OCAST Project HR4-064.]

11:15

4aPP10. A recovery model for coincidence neurons. Timothy A. Wilson, Rachod Thongprasirt (Dept. of Elec. Eng., Univ. of Memphis, Memphis, TN 38152), and Aziz A. Khanifar (Vanderbilt Univ., Nashville, TN 37240)

A new model for firings of coincidence neurons in the MSO is presented. The model is similar to one by Colburn *et al.* [Hearing Res. **49**, 335–346 (1990)]: Both are point processes, both have inputs which are

impulse-train representations of the firing events of primarylike neurons, both operate on a low-pass filtered sum of those inputs, and both output a sequence of impulses at the event times of the coincidence neuron's firings. In this model, their threshold-crossing neuron model is replaced by a recovery model similar to that modeling the input firings. The coincidence neuron's drive function is a nonlinearly distorted version of the filtered sum of inputs; for simplicity, its recovery function is the same as that of the input neurons. The model's firing rates, vector strengths, period histograms, and interval histograms are compared to data from dog and cat and with similar responses of the Colburn *et al.* model.

11:30

4aPP11. Neural responses to the onset of voicing are unrelated to other measures of temporal sensitivity. Donal G. Sinex and Guang-Di Chen (Arizona State Univ., Dept. of Speech and Hearing Sci., Tempe, AZ 85287-1908)

Voice onset time (VOT) is a temporal cue that can distinguish consonants such as /d/ from /t/. We have previously shown that neurons' responses to the onset of voicing are strongly dependent on their spectral sensitivity. This study examined the relation between neurons' temporal sensitivity, determined from responses to amplitude-modulated tones, and the same neurons' responses to the onset of voicing. Responses to VOT syllables and two types of modulated tones were obtained from low-frequency neurons in the inferior colliculus (IC) of the chinchilla. Both VOT and the modulation period varied from 10–70 ms in 10-ms steps. Neurons that respond selectively to modulated tones might be expected to respond strongly to the syllables whose VOTs match the preferred modulation periods. However, for most neurons the correlation between discharge rate to modulated tones and voicing onset was low. Some neurons exhibited moderate selectivity for certain modulation periods, but this selectivity was usually unrelated or even inversely related to the same neurons' selectivity for VOT syllables. Overall, responses to modulated tones did not account for the responses of IC neurons to the complex temporal cues associated with the onset of voicing. [Work supported by NIDCD.]

11:45

4aPP12. Regular interval stimuli: Are higher-order intervals necessary? William A. Yost and William P. Shofner (Parmly Hearing Inst., Loyola Univ. Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626)

Many stimuli that contain regular temporal intervals are perceived as having a pitch equal to the reciprocal of the interval. Autocorrelation is one method of determining the interval in these regular interval stimuli (RIS). Autocorrelation analysis has been used as a ''model'' for calculating the pitch and pitch strength of RIS. Autocorrelation is an ''all-interval'' analysis as opposed to the first-order interval analysis that is often used to describe the temporal properties of neural spike trains. Both actual neural responses (from the AVCN of the chinchilla) and those of a simple neural model were measured to determine the order of intervals contained within RIS. In a psychoacoustic experiment, RIS were generated with no first-order intervals or with only first-order intervals, and listeners discriminated between these stimuli and stimuli with no regular intervals. The results of the experiments and the calculations indicate that the pitch and pitch strength of RIS probably depend on only lower-order intervals (no higher than fourth-order intervals). These results will be discussed in terms of theories of pitch and timbre processing. [Work supported by NIDCD, DC-000293.]

THURSDAY MORNING, 4 DECEMBER 1997

SAN DIEGO/GOLDEN WEST ROOMS, 8:30 A.M. TO 12:00 NOON

Session 4aSC

Speech Communication: Vocal Tract Modeling, Speech Technology and Other Topics (Poster Session)

Anders Lofqvist, Chair

Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

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. To allow for extended viewing time, posters will remain on display until 5:00 p.m.

Contributed Papers

4aSC1. A finite-element template for tongue modeling. Reiner Wilhelms-Tricarico (Res. Lab. of Electron., MIT, Cambridge, MA 02139) and Chao-Min Wu (Univ. of Wisconsin, Madison, WI 53705)

This report describes the design, data structure, and applicability of a refined finite-element template for the human tongue and connected oral structures. The finite-element model relies mainly on data sets from the Visible Human Project of the National Library of Medicine. The model is composed of macroblocks, which represent geometric subsections of the tongue. These blocks represent, in some cases, either individually or in combination, functional subsections, such as individual muscles. For each macroblock (or geometric region) a finite-element mesh can be generated such that the whole of the tongue can be modeled by a mesh of finite-elements, since the subdivision is compatible across block boundaries. The finite element template contains information about muscle tissue distribution of the tongue and velopharynx. Approximate muscle fiber directions are represented as direction fields. The model can be adapted to individual morphology if a set of morphological landmarks can be specified from measurements such as those from MRI data of an individual.

4aSC2. A study of speaker adaptation of continuous density hidden Markov models. Prabhu Raghavan and Chiwei Che (CAIP Ctr., Rutgers Univ., Piscataway, NJ 08855)

Considerable research in the area of speech recognition, in the past two decades, using HMMs has led to successful reduction in word error rates (WER). Owing to their statistical nature, HMMs require large amounts of data to get good estimates of the model parameters. It has been shown that

a speaker dependent (SD) model has about 2–3 times less WERs compared to a speaker independent (SI) model. But the heavy data requirement from one speaker is quite unreasonable. Speaker adaptation techniques mitigate this problem by the use of SI data to obtain robust estimates and then use small amounts of SD data to adjust these parameters so that they are optimal to that speaker (in some optimality criterionlike ML). In this paper, the different techniques of adaptation are studied. The salient methods, namely MLLR and MAP, of each type are discussed in detail. A method of constrained unsupervised adaptation, which discards parts of the original hypothesis that are deemed to be incorrect, is proposed. This decision is made on past experience obtained on training data. This decision making module uses a neural network and a CART tree.

4aSC3. Mapping of muscle anatomy on three-dimensional magnetic resonance images of the human tongue based on morphological landmark selection. Chao-Min Wu (Waisman Ctr., Univ. of Wisconsin, 1500 Highland Ave., Madison, WI 53705) and Reiner Wilhelms-Tricarico (MIT, Cambridge, MA 02139)

The purpose of this study is to incorporate accurate muscle morphology in a computational 3-D tongue model based on anatomical literature and MRI data. Previous work [C.-M. Wu and R. Wilhelms-Tricarico, J. Acoust. Soc. Am. **100**, 2659(A) (1996)] defined several landmarks within the 3-D space of the tongue tissue and developed a computational method to map Miyawaki's drawing of muscle fiber directions [K. Miyawaki, Ann. Bul. RILP, Univ. Tokyo **8**, 23–50 (1974)] onto the cross-sectional tongue images from the visible human data. In this study, the same technique is applied to identify a set of corresponding landmarks between Miyawaki's data and MRI scans of a subject's tongue forming a consistent articulatory posture. The geometrical mapping of these two data sets is achieved by the thin-plate spline mapping method. The result shows the sufficiency and robustness of the chosen landmark points to represent a general 3-D structure of the human tongue. This study is expected to provide a computational basis for quantitative comparison of MRI data across subjects and vowel types.

4aSC4. Articulatory based low bit-rate speech coding. Samir Chennoukh, Daniel Sinder, Gael Richard, and James Flanagan (Ctr. for Comput. Aids for Industrial Productivity, Rutgers Univ., P.O. Box 1390, Piscataway, NJ 08855-1390)

A voice mimic system has been designed to achieve an articulatory description of the speech signal based on an analytic description of the vocal tract (VT) parameters as a function of the first two formant frequencies. Articulatory based low bit-rate speech coding using the analytic mapping technique has been demonstrated. The natural speech input is analyzed to obtain the first two resonances of the VT, and the VT shape is estimated from the analytic acoustic-to-articulatory mapping. The shape is modeled using three parameters. Then, the articulatory parameters in addition to the frequency of the vocal cord vibration are transmitted. The receiver restores the VT shape from the articulatory parameters, computes the formant frequencies corresponding to the VT shape and synthesizes the output speech using a formant synthesizer. The bit-rate of the coding is determined from the parameter sampling rate and their digital representation. An intelligible output speech signal has been obtained for bit-rates as low as 624 bits/s using parameter sampling at 48 times/s, 3 bits for each of the VT parameters and 4 bits for the vocal cord vibration frequency. Coding at this bit-rate has been demonstrated on voiced sentences. [Research supported by ARPA DAST 63-93-C-0064.]

4aSC5. An aeroacoustic fricative source model. D. Sinder, M. Krane, and J. Flanagan (CAIP Ctr., Rutgers Univ., Piscataway, NJ 08855)

A model for fricative sound generation based on aeroacoustic theory is implemented in an articulatory synthesizer. The aeroacoustic model is based upon Howe's reformulation of the acoustic analogy, which allows the source to be specified in terms of the jet formed at a constriction, the shape of the vocal tract walls, and the acoustic loading due to the lip termination. As such, the model uses existing acoustic models for the vocal tract combined with an approximate model for the behavior of the jet based on turbulence theory and experiment. This allows the fricative source strength, impedance, and temporal characteristics to be specified using the lung pressure, the constriction diameter, and the shape of the vocal tract downstream of the constriction. Evaluation of the source model is presented. Computed estimates of fricative source characteristics are presented and compared to source models used by other investigators. A comparison of synthetic fricative sounds generated using these models is also presented. [Research supported by NSF/ARPA IRI-9314946 and ARPA DAST 63-93-C-0064.]

4aSC6. Labeling a speech database with landmarks and features. Jeung-Yoon Choi, Erika Chuang, David Gow, Katherine Kwong, Stefanie Shattuck-Hufnagel, Kenneth Stevens, and Yong Zhang (Res. Lab. of Electron., MIT, Cambridge, MA 02139)

A database of utterances has been labeled for acoustically salient landmarks and for phonetic features. The landmarks locate syllabic nuclei, glide midpoints, and acoustic discontinuities at the narrowing or releasing of consonantal constrictions. The consonantal landmarks are labeled with values of the articulator-free features (e.g., [continuant] and [sonorant]), and values of articulator-bound features are recorded at each landmark based on acoustic evidence near the landmark. These features specify, for example, the major articulator and its placement, as well as the activity of secondary articulators such as the glottis and the soft palate. Conventions have been developed for labeling reduced vowels, syllabic consonants, glottal stops, consonant sequences in which closures or releases are not acoustically marked, etc. The labeled database is being used for two purposes: (1) to provide a documentation of instances in which the acoustic/articulatory representation of an utterance deviates from the underlying feature-based lexical representation and (2) to train and to test signal-processing algorithms that are being developed to locate the landmarks and to identify the features attached to these landmarks. An advantage of this labeling method is that it gives a more fine-grained phonologically-relevant representation of an utterance than a conventional labeling in terms of phonetic symbols. [Supported in part by NIH Grant No. DC02978.]

4aSC7. New parameters and mapping relations for the HLsyn speech synthesizer. Helen M. Hanson, Kenneth N. Stevens, and Robert E. Beaudoin (Sensimetrics Corp., 126 Landsdowne St., Cambridge, MA 02139)

HLsyn is a quasiarticulatory speech synthesizer in which a small set of parameters control a Klatt synthesizer [Stevens and Bickley, J. Phon. **19** (1991)]. Originally, ten parameters were used. In this paper three new physiologically based parameters, together with some additional modifications, are described. The first is a time-varying subglottal pressure parameter, which provides the user with additional control of the voice-source amplitude. It can also be used to turn on voicing, and it has an influence on fundamental frequency. The second parameter is a time-varying percentage change in the compliances of the vocal-tract walls and vocal folds. This parameter can, for example, be employed when synthesizing voiced obstruents: increasing it results in facilitation of glottal vibration and lowering of the fundamental frequency. The third parameter is the time-varying cross-sectional area of a posterior glottal chink. Because it is independent of the area at the membranous folds, significant noise-source amplitudes can be achieved during voiced speech. Thus one can synthesize aspirated voiced stops, as well as breathy speech, in a natural manner. Finally, a default female voice and an intrinsic pitch feature have been added to the system. Examples of copy synthesis where the new parameters play a role are presented. [Work supported by NIH Grant MH52358.]

4aSC8. An HMM-based measure for evaluating the quality of foreign-accented speech. Robert Smith, Kiyoaki Aikawa, and Satoshi Takahashi (NTT Human Interface Labs, 1-1 Hikarino-oka, Yokosuka-shi, Kanagawa 239, Japan)

A new criterion for measuring the quality of foreign-accented speech is proposed. The proposed criterion uses both the absolute and relative quality scores produced by an HMM speech recognizer. An absolute scoring function is determined by the difference of likelihood distributions between the correct phone HMM and an incorrect phone HMM for a speech segment. A relative scoring function is established to discriminate a delta-self value distribution from a delta-sub value distribution. A delta-self value is obtained by the likelihood score difference between the correct model and an incorrect model for a speech segment. A delta-sub value is obtained by the likelihood score difference between two incorrect phone models against the speech segment. The new idea is motivated by the following evidences. It is sometimes found that the absolute score is low, even if the speech segment sounds correct. In that case, however, the relative score is still high for the speech segment. This may happen because the HMM parameters were not chosen for evaluating voice quality. The best speech quality meets a condition that both the absolute and the relative scores are high. A practical integrated measure is the linear combination of the absolute and the relative scoring functions.

4aSC9. Indexical cuing in laughter. Jo-Anne Bachorowski, Moria J. Smoski (Dept. of Psych., 301 Wilson Hall, Vanderbilt Univ., Nashville, TN 37240, bachja@ctrvax.vanderbilt.edu), and Michael J. Owren (Cornell Univ., Ithaca, NY 14853)

Laughter is a frequently produced yet poorly understood vocal signal. Despite the lack of basic knowledge regarding this species-typical vocalization, this signal may be advantageous to study as it provides an opportunity to examine indexical cueing in the absence of linguistic content. The source-filter model of vowel production guided acoustic analysis of laughter recorded from experimentally naive subjects as they watched emotion-inducing film clips in a laboratory setting. Subjects were 48 males and 48 females tested either alone or with a same- or opposite-sex friend or stranger. Laugh acoustics were found to vary by gender, talker, and social context. Males produced predominantly unvoiced, ''grunt'' laughs, whereas female laughter was more variable and included more voiced, long duration ''song'' laughs. Discriminant function analyses revealed that source and filter-related cues were differentially associated with social context and talker identity. Variability in source-related cues, particularly $F0$, more successfully discriminated social context, whereas cues related to the vocal tract transfer function better discriminated among individuals. Females, but not males, had significantly higher formant frequencies when tested with opposite-sex strangers. The overall pattern of results is consistent with recent examinations of indexical cuing in human speech and nonhuman primate vocalizations. [Work supported by NIMH.]

4aSC10. Parsing the contribution of lip and tongue position to fricative noise frequency. D. H. Whalen and Bryan Gick (Haskins Labs., 270 Crown St., New Haven, CT 06511)

The English alveolar and palatal fricatives have, as part of their realization, high-frequency noises that are shaped both by the rounding and protrusion of the lips and the position of the tongue on the hard palate. The acoustic effects are of two main types, influence on the main spectral resonance of the noise, and the frequency and intensity of a second formant resonance within the noise. However, the exact contribution of each of these articulatory factors to the two acoustic effects is not known. The present study used electromagnetometer measures of tongue and lip location and video-based measures of lip aperture, area, and protrusion. Initial data indicate that area of the lip opening correlates significantly with $F2$ resonance, while tongue position correlates with fricative pole. Surprisingly, upper lip protrusion correlates with both the $F2$ resonance and the fricative poles. These preliminary results suggest that both lip and tongue configuration contribute to the shape of the fricative poles, while the $F2$ resonance is primarily dependent on the lip configuration. Implications for perceptual parsing will be discussed. [Work supported by NIH grant DC-02717.]

4aSC11. Predicting pharynx shape from the tongue during vowel production. Min Kang (Haskins Labs., 270 Crown St., New Haven, CT 06511 and Dept. of Linguist., Yale Univ., New Haven, CT 06520), D. H. Whalen and Harriet Magen (Haskins Labs., New Haven, CT 06511),

The shape of the pharynx greatly affects the formants of vowels, yet the measurement of pharyngeal shape is relatively difficult. Using midsagittal MRI images of a male speaker producing eleven English vowels, the predictability of pharyngeal shape during vowel production was assessed. There were two categories of predictors: measured and categorical. Measured predictors were four points on the tongue representing the points likely to be locations for receiver coils in an electromagnetometer system. Categorical predictors were vowel frontness and height, and the typical position of the epiglottis relative to the base of the tongue (as seen in MRI images). Midsagittal distance was measured at 3-mm intervals along the vocal tract length. A stepwise regression using measured and categorical variables gave good predictions for all distances along the vocal tract (r greater than 0.80), except near the uvula. (Using just the categorical predictors resulted in surprisingly good estimations.) Position of the uvula varied across repetitions of the vowels, and so exclusion of the uvula increased predictability in the uvular region. For static vowel patterns in English, then, pharynx shape is fairly predictable from tongue shape. [Supported by NIH grant DC-02717.]

4aSC12. Name retrieval from pronunciation and spelling over the telephone, using HMM modeling and robust directory access. G. Gravier, F. Yvon, G. Etorre, and G. Chollet (ENST, 46 rue Barrault, 75634 Paris Cedex 13, France)

A system to retrieve names in a directory from their pronunciation and their spelling is presented, for telephone quality speech. Most of the previous approaches to letter recognition for spelled names are knowledge-based because of the high phonetic confusability between letters [C. S. Myers and L. R. Rabiner, J. Acoust. Soc. Am. **71**(3), 716–727 (1982)] [Cole *et al.*, Eurospeech'91, 479–482]. To avoid the knowledge-based approach, HMM modeling of speech units (SU) is used with word models for letters and phone models for name pronunciation. The directory is built automatically from a list of names using a grapheme-to-phoneme converter for the name and rules for the spelling, each entry in the directory consisting of several variants for both the pronounced and spelled names. From the acoustic recognition, the corresponding entry in the directory is found using DTW or a tree-search algorithm. Both methods allow insertion and deletion, and the cost for a substitution can either be fixed or determined from the train corpus confusion matrix. On a test database of 5000 names uttered by 3000 speakers ($\sim$ 50 000 SU), 70% of the units are correctly recognized. Preliminary experiments based on spelled names with a DTW lexical search, showed that substitution costs defined using the confusion matrix significantly improved the results.

4aSC13. Stochastic linear regression analysis of speech production data. H. Betty Kollia (Dept. of Logotherapy, TEI Patras, Patra, Greece) and Jay Jorgenson (Dept. of Math., Oklahoma State Univ., Stillwater, OK 74074)

Dynamic models in speech production have been traditionally based on linear regression analysis of articulatory data. The purpose of this work is to introduce a stochastic calculus methodology to the analysis of such

data in speech production models. The issue of variability encountered in the analysis and description of articulatory movements during speech may be addressed with the implementation of a stochastic term, thus placing the variability in a controlled probabilistic framework. Upper lip, lower lip, jaw, and velum movements tracked optoelectronically at Haskins Laboratories [Kollia *et al.* J. Acoust. Soc. Am. **98**, 1313–1327 (1995)] during production of a test utterance are described in this manner. The authors believe that this algorithm has wide applicability since it yields confidence interval rather than point value estimates for kinematic data description. [Work supported by NSF.]

4aSC14. Improvement in quality of speech received by a mistuned receiver at suppressed carrier SSB. Shin'ichiroh Kaneko, Jouji Suzuki, and Tetsuya Sjimamura (Dept. of Information and Comput. Sci., Saitama Univ., Shimo-okubo 255 Urawa, 338 Japan)

Single sideband modulation (SSB) is the most efficient speech transmission system from the viewpoint of power consumption and bandwidth occupation. It is widely used in amateur radio, mobile radio, and others. At suppressed carrier SSB(SCSSB), the receiver should tune to the carrier frequency accurately. Manual tuning, however, is very difficult due to the lack of the carrier signal. It fails in fine tuning at the reception and the quality of received speech degrades seriously owing to the shift of speech spectrum. This paper proposes a method to improve the quality of speech degraded at SCSSB. Received speech is transformed into the log spectrum and its autocorrelation function (ACLOS) is calculated. If tuning is carried out correctly, the log spectrum of voiced sound has a harmonic structure. Mistuning gives a spectrum shift on that frequency axis. On the other hand, ACLOS is always a periodic signal that corresponds to the harmonic structure. By calculating a cross-correlation function between ACLOS and the log spectrum, mistuning frequency can be easily estimated. This method is able to estimate the mistuning frequency within the accuracy of 10 Hz, even in noisy channels. After obtaining the mistuning frequency, the shifting spectrum of received speech is carried out and more natural speech is obtained.

4aSC15. The effects of aging and rate on tongue shape variability and on preservation of position in glides. Maryam Jaberi, Maureen Stone (Dept. of Otolaryngol., Univ. of Maryland Med. School, 16 S. Eutaw St., Ste. 500, Baltimore, MD 21201, mstone@umabnet.ab.umd.edu), Suzanne Gray (Johns Hopkins Univ., Baltimore, MD 21218), and Elsbieta Slawinski (Univ. of Calgary, Calgary, Canada)

Slowing of neural processes due to aging may cause effects in motor speech production. In a study of /w/, Slawinski [E. B. Slawinski, J. Acoust. Soc. Am. **95**, 2221–2230 (1994)] found a slower prevoicing stage, and a smaller off-glide frequency range for older speakers than younger. The latter finding suggests that older speakers centralize their articulatory gestures. Articulatory gestures may also centralize at faster speech rates. The present study compared midsagittal tongue gestures for /w/, for older versus younger subjects, matched for rate of speech. Ten female subjects (five aged 21–29, five aged 64–71) repeated: /əwVp/ and /əgwVp/ using the vowels /i,a,u/, five times each, while ultrasound video recordings were made. The hypothesis was that coarticulationwould cause different /w/ positions, (higher in /g/ context). If older subjects centralize tongue position for /w/, they might produce an even greater difference, if the high /g/ tongue position facilitated a more extreme /w/. Target frames were digitized and tongue surface edges were extracted [M. Unser and M. Stone, J. Acoust. Soc. Am. **91**, 3001−3007 (1992)]. The distances between /w/ and /(g)w/ were determined using $L2$ norms. For two older and two younger speakers, rate effects were observed and age effects were not.

4aSC16. An automatic speech recognition system using time-delays self-organizing maps with physiological parametric extraction. Jose M. Ferrandez, Daniel del Valle, Victoria Rodellar, and Pedro Gomez (Laboratorio de Comuncacion Oral R. W. Newcomb, Facultad de Informatica, Univ. Politecnica Madrid, C. Montegancedo, Boadilla del Monte, 28660 Madrid, Spain, jmanuel@naranjo.datsi.fi.upm.es)

Physiological parametric extraction uses auditory models as a front end for speech recognition. These last methods assume that if speech signals are coded in the same way that the auditory system does, speech could be later identified showing the main properties that biological systems do: robustness and accuracy. The proposed system consists of a cochlear model implemented by gammatone filterbanks as proposed by Patterson [J. Acoust. Soc. Am. **96**, 1409–1418 (1994)]. This stage will feed a nonlinear mechanical-to-neural transduction module based on the Meddis hair-cell model [J. Acoust. Soc. Am. **79**, 702–711 (1986)], which will compute auditory-nerve firings. Finally, a temporal integration/component–extraction module will integrate neural patterns for identifying the relevant components embedded in the speech signals [characteristic frequency (CF), frequency modulation (FM) and noise burst (NB)], which are shared by human speech and animal sounds for communication. The model adopts a spatiotemporal strategy, which uses temporal information in low CF fibers (phase-locking mechanism) and spatial information for the higher ones. The recognizing module consists in a time-delay self-organizing map, which will capture not only the spectral variability contained in the signal, but also the temporal one, providing better generalization properties. [Work supported by NATO CRG-960053.]

4aSC17. Crime automatic speaker verification and identification system. Igor I. Gorban (Inst. of Math. Machines and Systems, Acad. Glushkov, 42 Ave., Kiev, Ukraine 252187, gorban@olinet.isf.kiev.ua)

A new automatic speaker verification process was presented at the Eighth Annual International Conference on Signal Processing Applications and Technology (ICSPAT '97), San Francisco, CA (I. I. Gorban, Robust Speaker Verification Algorithms). On the basis of these algorithms the crime automatic speaker verification and identification (CASVI) system has been created. The main feature of the system is that it is possible to verify and identify a speaker when the speech records are very corrupted and/or contain extraneous noise. Testing conducted under different conditions has shown that the CASVI system works well when the level of frequency distortions is lower than 35 dB and the level of the signal-to-noise ratio is lower than 12 dB. These results have been obtained for a new modification of the algorithms in which the adaptive threshold verification process is included and the identification algorithms are added. A description of the CASVI system, it's algorithms, the methodology, and the results of the system testing will also be presented at the meeting.

4aSC18. A dynamic neural network model of speech production in the developing child. Daniel E. Callan, Ray D. Kent, and Houri K. Vorperian (Dept. of Commun. Disord., Univ. of Wisconsin, Madison, WI 53706)

Many theories of speech processing propose the existence of motor control systems that utilize invariant vocal tract configurations to specify particular goals of speech production. One problem that challenges these theories is the fact that the associated structures involved with speech production go through a considerable amount of change during development. The same speech goals continue to be achieved during the course of development despite the changes that the vocal tract configuration undergoes. In this paper, speech production in the developing child is modeled by a dynamic neural network that incorporates value-dependent learning based on self-produced auditory stimulation across self-organizing sensorimotor maps. The conversion from articulatory configuration to acoustic signal is worked out by a modified version of the Maeda articulatory model that utilizes developmental parameters. The performance of the neural network was assessed at different points during development by determining the articulation-acoustic output needed to produce various

segments of speech in relation to the corresponding trajectory of activation across the neural network. The results are discussed in relation to a dynamic theory of speech processing that can account for developmental change as well as some speech disorders. [Work supported by NIDCD.]

4aSC19. An acoustic model of gesture overlap: Further studies. Gary Weismer and Jeff Berry (Dept. Commun. Disord. and Waisman Ctr., 1500 Highland Ave., Univ. of Wisconsin, Madison, WI 53705)

Previous work from our laboratory has suggested that certain acoustic measures may serve as indices of the extent of overlap for successive articulatory gestures. These experiments have employed speaking-rate variation as a means to vary the extent of articulatory gesture overlap, in a number of different syllable types. Explicit models have been used to predict how rate-induced variation in gesture overlap would be realized in the acoustic output of the vocal tract. Specifically, the $F2$ onset and offset values at CV and VC interfaces, respectively, where C and V are constant, have been shown to be sensitive to the rate variable, sometimes approaching the magnitude of variation associated with CV syllables where C is constant and V is varied (i.e., the classic vowel-induced coarticulation paradigm). However, the results to date appear to be dependent on specific syllables, but the reasons for this specificity are unknown. In the present study, some systematic analyses of vowel effects on the acoustic model of gesture overlap are presented. Results will be discussed in terms of the goodness of the model and theories of coarticulation.

4aSC20. Evaluation of a reiterant force impulse task: A constructive replication. Kate Bunton and Gary Weismer (Dept. of Commun. Disord., Univ. of Wisconsin, 1500 Highland Ave., Madison, WI 53705)

In the current study, characteristics of labial force–impulse tasks were investigated under three speaking conditions using a minipressure transducer. In the task, neurologically normal adults were first asked to read aloud two sentences of different lengths (ten repetitions each) containing bilabial voiced and unvoiced consonants, and were then required to repeat the series of utterances using mimed speech. Mimed speech was produced with subjects holding their breath; subsequently, the experimenter cued subjects to use a normal airflow and the mimed utterances were repeated. Subjects were then trained to produced sequences of labial force impulses modeled on the real-speech utterances by substituting /ba/ or /pa/ for each syllable. The reiterant condition was also recorded as mimed. The goal of the analyses were to (a) compare the timing of the real-speech sequences to the timing of mimed and reiterant versions, and (b) compare the force magnitudes associated with labial stress produced during real speech to mimed and reiterant repetitions. Results will be discussed in terms of the relationship between orofacial, nonspeech motor performance and speech production performance. [Work supported by NIH Award No. T32 DC00042 and R01 DC00319.]

4aSC21. Magnetometer observation of articulation in sitting and supine conditions. Mark K. Tiede, Shinobu Masaki, Masahiko Wakumoto, and Eric Vatikiotis-Bateson (ATR Human Information Processing Res. Labs., 2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan, tiede@hip.atr.co.jp)

Magnetic resonance imaging (MRI) has recently emerged as a useful tool for the investigation of the three-dimensional vocal tract shape in sustainable articulatory targets such as vowels and fricatives. However, existing scanning devices require that subjects phonate in a supine position. Previous work [Whalen, J. Acoust. Soc. Am. Suppl. 1 **88** (1990)] compared MRI to cineradiographic data collected in a sitting posture, and reported limited effects of gravity on velar height differences in vowel production. This study used electromagnetometry [Perkell *et al.*, J. Acoust. Soc. Am. **92**, 3078–3096 (1992)] to investigate whether detectable systematic differences in midsagittal tongue posture exist between the two phonation conditions. Subjects fitted with nine transducers (four on the tongue, mandible, upper and lower lips, and two for head correction reference) produced both running speech and sustained vowels mimicking MRI protocol, repeated in sitting and supine postures. Analysis of separately recorded acoustics showed systematic effects of production posture on lower formant bandwidth. Results of comparing tongue height differences between the two conditions will be presented.

4aSC22. MRI observation of dynamic articulatory movements using a synchronized sampling method. Shinobu Masaki, Mark K. Tiede, Kiyoshi Honda (ATR Human Information Processing Res. Labs., 2-2 Hikaridai Seika-cho Soraku-gun, Kyoto 619-02, Japan, masaki@hip.atr.co.jp), Yasuhiro Shimada, Ichiro Fujimoto, Yuji Nakamura (Takanohara Chuo Hospital) and Noboru Ninomiya (Shimadzu Corp.)

The use of magnetic resonance imaging (MRI) in speech research has, typically, been limited to static, sustainable vocal tract configurations due to the long acquisition times required [e.g., Baer *et al.*, J. Acoust. Soc. Am. **90**, 799–828 (1991)]. This work describes the application of a synchronized MRI sampling technique developed for cardiac motion imaging to the visualization of dynamic articulatory movements. Subjects produced 128 repetitions of a target utterance (/tata/ and /kaka/) coincident with audible tone bursts synchronized with gating pulses controlling MRI scanning. During each repetition, 35 scans at 25-ms intervals were acquired, each scan providing one row of frequency-domain data for each of the 35 output frames. After repetitive scans were completed, the full sequence was processed to give a 40-frames-per-second midsagittally oriented movie. Analysis of the resulting movies revealed that tongue shapes for the interconsonantal /a/ varied between the two utterance types, indicating a place-of-articulation effect of consonants on coproduced vowel tongue shape. This suggests that the synchronized sampling method provides a practical technique for investigation of the dynamic characteristics of articulatory movements during short utterances.

4aSC23. Three-dimensional visualization of electropalatography. Masahiko Wakumoto and Shinobu Masaki (ATR Human Information Processing Res. Lab., 2-2 Hikari-dai, Seika-cho, Soraku-gun, Kyoto, 619-02 Japan, mwakumo@hip.atr.co.jp)

Electropalatography (EPG) is a widely used noninvasive technique for monitoring dynamic tongue-palate contact patterns in speech research and speech clinics. However, the conventional two-dimensional (2D) visualization system is incapable of informing individual characteristics of three-dimensional (3-D) palatal shape or the relationship between electrode positions and anatomical landmarks (e.g., teeth). This paper reports a method for visualizing tongue-palate contact patterns on the 3-D palatal shape display. A convex palatal shape model was formed by silicone impression material from an individual dental cast, and electrode positions on the individual-specific artificial palate (RION SP-01, 63 electrodes) were marked on the surface of the convex model. The 3-D geometry and the surface texture information of the convex model were extracted by a laser beam scanner (Cyberware 4020). Then, palatographic data were collected through the EPG system (RION DP-20) and the tongue-palate contact patterns were mapped on the 3-D graphical display of the palate. Preliminary observation of Japanese fricatives revealed vertical differences in constriction between [s] and [ʃ]. Its application to speech production research and speech training will be discussed.

4aSC24. Geometrical display of speech spectra as an aid to lipreading. Eric J. Hunter and William J. Strong (Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602)

A geometrical display of speech spectra intended as an adjunct to lipreading was developed. Spectra were calculated at 5-ms intervals from speech sound pairs ambiguous to lipreaders. The spectra were displayed as sequences of irregular decagons. Human subjects were asked to discrimi-

nate between pairs of spectral decagon sequences derived from pairs of ambiguous speech sounds. Subjects were able to discriminate between most of the visual spectral patterns derived from ambiguous sounds. However, spectral patterns associated with the voiced/unvoiced contrast in some stop pairs were not discriminated consistently.

4aSC25. Interarticulator programming in VCV sequences. Anders Lofqvist and Vincent L. Gracco (Haskins Labs., 270 Crown St., New Haven, CT 06511)

This study examined the coordination of tongue body and lip movements in VCV sequences, where the consonant is a bilabial stop /p, b/, and the vowel one of /i, a, u/; only asymmetrical vowel contexts were analyzed. Lip and tongue timing and vowel-to-vowel tongue body kinematics were examined. The movements were recorded using a magnetometer system. Four subjects participated and produced ten tokens of each sequence. The phasing between the onset of the lip closing movement and the onset of the tongue body movement from the first to the second vowel varied across subjects, with no obvious pattern according to consonant voicing or vowel context. Overall, the onset of the tongue movement to the second vowel started earlier relative to the lip closing movement as the length of the tongue movement trajectory increased. During the oral closure, 40%–80% of the tongue movement path was traversed, again with no clear pattern across subjects. Consonant voicing reliably influenced the tongue movement from the consonant release to the second vowel. During this interval, the tongue traversed a longer path for voiceless than for voiced stops, reflecting the longer voice onset time for the voiceless stops. [Work supported by NIH.]

4aSC26. Facial animation using visual polyphones. Åsa Hällgren and Bertil Lyberg (Telia Res. AB, Spoken Lang. Processing, S-136 80 Haninge, Sweden)

Today, synthetic speech is often based on concatenation of natural speech such as polyphones. So far there has mainly been two ways of adding a visual modality to such a synthesis: Morphing between single images or concatenating video sequences. In this study, however, a new method will be presented where recorded natural movements of points in the face are used to control an animated face. During the recording of polyphones to be used for synthesis, the movements of a set of markers placed in the face of the speaker are simultaneously registered in three dimensions. Connected to each acoustic polyphone will be a perfectly synchronized set of dynamic movement patterns, a visual polyphone. These visual polyphones are then concatenated in the same manner as the acoustic polyphones, and used to control the movements of an animated face. In the presented system, demisyllables are used as units both for the acoustic and the visual polyphones and it seems as these units will, in a good way, reflect the coarticulation process in natural speech.

4aSC27. On the meaning and the accuracy of the pressure-flow technique to determine constriction areas within the vocal tract. Xavier Pelorson, Salima Fahas, and Pierre Badin (Institut de la Commun. Parlée, INPG-Univ. Stendhal, 46 Ave. Félix Viallet, 38031 Grenoble Cedex, France, pelorson@icp.grenet.fr)

Since the work of D. W. Warren and A. B. Dubois [Cleft Palate J. **1**, 52–71 (1964)] was published, the ''pressure-flow technique'' has been widely used to estimate constriction areas within the vocal tract. However, two fundamental questions regarding this technique have not been clearly addressed: (1) What is exactly measured (minimum, maximum, or ''mean'' areas)? (2) Which accuracy can be expected from this technique? A theoretical and experimental study based on a vocal tract mechanical model, including various constriction shapes, is presented. Compared with much more complex viscous flow solutions, a simple one-dimensional flow model is shown to yield fair estimates of the areas (within 10%), except for low Reynolds numbers. The empirical head-loss factor $K=0.67$, sometimes used, appears meaningless and is probably due to an experimental artefact. The pressure-flow technique is shown to be relatively insensitive to the exact constriction shape (converging, straight, or diverging), and the estimated area to be close to the minimum area of the constriction. This result can be theoretically rationalized by considering that in all cases studied here the flow separation point is always close to the minimum constriction. Extensions and limits of these conclusions for unsteady flows will also be discussed.

4aSC28. An experimental study of the open-end correction coefficient for side branches within an acoustic tube and its application in speech production. J. Dang (ATR HIP, 2-2 Hikaridai Seikacho Soraku-gun Kyoto, 619-02 Japan), C. Shadle (Univ. of Southampton, UK), and H. Suzuki (College of Sci. and Technol., Tohoku, Japan)

The open-end correction coefficient (OECC) for side branches within an acoustic tube was investigated experimentally, and combined with morphological data of the vocal tract to arrive at OECC estimates useful in speech production. A number of mechanical acoustic models were used to examine the effects of the angle of the branch axis, and the proximity of the main tube around the open end of the branch. The results indicate that the OECC (α_1) depends on the cross-dimension L of the main tract at the branching point and the branch diameter D ($\alpha_1=0.414+0.517\cdot\log_{10}(L/D^2)$, $0.158<L/D^2<6.97$). For side branches connected to the main tract by a narrow neck, the OECC of each end of the neck is determined by $\alpha_2=0.82(1-\xi)/18\xi^2-3.1\xi+1$, where $\xi=$ the ratio of radius of the neck to that of the adjacent section. Using morphological data of the vocal-tract, the range of appropriate OECC values for estimating accurate vocal-tract transfer function was evaluated. The OECC of the nasopharyngeal port was about 0.8 for the oral-tract side, 0.3–0.6 for the nasal-tract side. For the ostia of the paranasal cavities, the value was ~ 0.8 for the cavity side, ~ 0 for the nasal-tract side.

4aSC29. A physiological model of the tongue and jaw for simulating deformation in the midsagittal and parasagittal planes. Jianwu Dang and Kiyoshi Honda (ATR Human Information Res. Labs., 2-2 Hikaridai Seikacho Soraku-gun Kyoto, 619-02 Japan)

A physiological articulatory model was developed for simulation of 3-D deformation and movement of speech organs. The model is composed of a midsagittal portion of the tongue and the jaw and hyoid bone that yield rotation and translation. The tongue model consists of a midsagittal layer and two parasagittal layers with 1-cm intervals. The form and muscle geometry of the whole model are extracted from MR images of a speaker. All the soft tissue and rigid organs are modeled by node points with mass and two types of branches to adjacent points: soft branches for muscles and connective tissue between node points, and hard branches for bony organs. The muscles are defined based on an improved Hill's model. Muscle activation signals, in part derived from EMG recordings, serve as the model's control parameters. Preliminary experiments revealed that the model can produce realistic 3-D deformation and movement for vowels and a sustained consonant /r/. Since all the displacements of soft tissue and rigid organs are computed with the same algorithm, the model produces plausible dynamic patterns of the tongue–jaw complex with relatively short computational time in comparison with the finite-element method.

Session 4aSP

Signal Processing in Acoustics: General Signal Processing

Brett Castile, Chair
Orincon Corporation, 9363 Towne Centre Drive, San Diego, California 92121

Contributed Papers

9:00

4aSP1. An application of laser-based ultrasonic nondestructive evaluation using a fiber-optics-based Fabry–Perot interferometer. Robert D. Huber, James V. Candy, Diane J. Chinn, and Graham H. Thomas (Lawrence Livermore Natl. Lab. P.O. Box 808, L-333, Livermore, CA 94550)

Fiber optics lend increased flexibility to laser-based ultrasonic nondestructive evaluation (NDE). In this work, fiber-optic cables are used to transmit light from a laser to the detection site, and then from the detection site to a Fabry–Perot interferometer. The use of fibers allows both the detection laser and interferometer to be placed at a considerable distance from the object under test. A direct line-of-sight of the object from the main equipment is not required, since the fibers may be fed through walls and around obstacles. In addition, by containing the laser light in the fibers, the chance of accidental exposure to powerful laser beams that may otherwise be transmitted through air is decreased. Laser-based ultrasonics is generally less sensitive to traditional contact ultrasonics, and in addition, some light is lost in the coupling of laser light energy into optical fibers, further decreasing the sensitivity; thus the need for signal processing of the received signals is of great importance. In this work, the waveforms obtained using the Fabry–Perot interferometer and the corresponding signal processing performed on the data to enhance the resulting image for NDE are discussed.

9:15

4aSP2. On-line measurement of error-path impulse response for an adaptive ANR headset. George J. Pan and Anthony J. Brammer (Inst. for Microstruct. Sci., Natl. Res. Council, Montreal Rd., Ottawa, ON K1A 0R6, Canada)

A digital, adaptive ANR headset system has been developed and tested in the laboratory, using a filtered-X LMS adaptive control algorithm and a floating-point digital signal processor (DSP). The implementation of the filtered-X LMS adaptation algorithm requires an impulse response (IR) model of the so-called ''error-path,'' which consists of the serial connection of an earphone, earcup acoustics, microphone, and electronic amplifiers and filters. Though the IR model can be measured off-line for laboratory studies using standard measurement procedure, on-line measurement or identification, using the DSP engine and digital signal processing techniques, is considered necessary and advantageous for practical applications. This paper discusses a DSP-based identification technique for the error-path IR model, which employs the well-known pseudorandom binary signal (PRBS), also called the maximum length sequence (MLS) signal, as the input to the error path. In this paper, the measurement method and system will be discussed. As an example, the error-path impulse response and the corresponding ANR results obtained in a laboratory setup will be presented. [Work supported by DCIEM, Toronto, Canada.]

9:30

4aSP3. Experiments on bit-rate reduction for orthophony signals. Ana M. Monsalve Caballero (Inst. of KW, Tech. Univ. of Berlin, Einsteinufer 17, Sekt. EN 8, R 452, 10587 Berlin, Germany)

Orthophony is the art of recording an acoustical sound field from a room and the reproduction of it by means of loudspeakers. For the reproduction of the signals, the simplest method is to use a transference matrix with the signals. The problem of the matrices is the sizes that they have, because they usually present dimensions too difficult to transmit or/and store. A way to solve the problem is using bit-rate reduction. Based on orthophony signals recorded in the studio of the institute of KW (Kommunikationswissenschaft) of the Technical University in Berlin and their corresponding matrices, experiments were accomplished for the reduction using MATLAB since it facilitates especially the use of matrices. The behavior of the different methods with different signal types (for example, speech, music, and noise) was observed and analyzed. The results are presented in this work. [This work was supported by Professor M. Krause, Institute of KW. Tech. Univ. of Berlin, Germany.]

9:45

4aSP4. Voiced excitation functions calculated from micropower impulse radar information. G. C. Burnett, T. J. Gable, J. F. Holzrichter, and L. C. Ng (Lawrence Livermore Natl. Lab., L-437, P.O. Box 808, Livermore, CA 94551, burnett5@llnl.gov)

Efforts underway at the Lawrence Livermore National Laboratory to use newly designed micropower impulse radars (MIR) to measure in real time the excitation function of the vocal tract will be presented. Studies undertaken in collaboration with the University of California at Davis and the University of Iowa with high-speed laryngoscopic cameras, electroglottographs, flow masks, and subglottal pressure transducers have solidified the relationship between the signal returned by the MIR and the voiced excitation function of the vocal tract. As a result, for the first time a transfer function of the vocal tract can be calculated in real time and with unprecedented clarity for voiced speech. This new capability could have significant implications for improvements in speech recognition and speech synthesis processing.

10:00

4aSP5. Comparison of conventional acoustic and MIR radar/acoustic processing of speech signals. T. G. Gable, G. C. Burnett, J. F. Holzrichter, L. C. Ng (LLNL, P.O. Box 808, L437, Livermore, CA 94551, gable2@llnl.gov), and W. A. Lea (Speech Sci. Inst., Apple Valley, MN 55124-0428)

Applications of the micropower impulse radar (MIR) to speech research at the Lawrence Livermore National Laboratory has produced potentially new methods of speech processing. They include the accurate calculation of vocal tract transfer functions, formant, and pitch analysis, and basic phoneme synthesis. These speech parameters have traditionally been in the realm of all-pole LPC calculations. Related research using the MIR radar has supplied an increasingly accurate voiced excitation function, which makes possible transfer function calculations using both poles

and zeros, yielding more accurate formant information and more natural sounding synthesis. This paper compares the newly obtained results with traditional LPC and cepstral approaches and demonstrates the improvements based on experimental data from several male and female subjects. The radar data also allow extremely accurate pitch tracking, which is simpler and more robust than that calculated by traditional means. This information can significantly enhance acoustic-based speech processing by allowing pitch adaptive processing.

10:15

4aSP6. Measurements of normal incidence surface impedance using the single transfer function method with multitone periodic random noise signal. Xavier Olny and Franck C. Sgard (LASH, DGCB URA CNRS 1652, ENTPE—Rue Maurice Audin, 69518 Vaulx-en-Velin Cedex, France, xavier.olny@entpe.fr)

Commonly used signals for determining the transfer function of a linear filter are pure tone excitation, white noise, or a pseudorandom noise (MLS) signal. However, it is known that the accuracy of this measurement depends strongly on the kind of signal used. In order to try to improve the quality of the measurement, multitone periodic random noise signal properties are investigated in the case of the normal incidence surface impedance measurement. Kundt's tubes and free-field impedance results using a single microphone transfer function method are given in a large frequency range (20–4000 Hz) for a multitone periodic random noise signal. These results are compared with those obtained with pure tone, white noise, and MLS signal excitation. Also, the influence of noise immunity is investigated for each type of signal.

10:30

4aSP7. Plane wave reflection coefficient estimation by use of spatial parametric signal modeling. Hyu-Sang Kwon and J. Stuart Bolton (1077 Ray W. Herrick Lab., School of Mech. Eng., Purdue Univ., West Lafayette, IN 47907-1077)

The plane wave reflection coefficient of an absorbing infinite plane can be estimated from pressure distributions on two lines parallel to the surface. The measured pressures are decomposed into incoming and outgoing plane waves in the wave-number domain after the Hankel transform is applied to the spatial data. In practice, only a finite number of pressure measurements can be made on the measurement lines. As a result, errors resulting from aliasing and truncation are induced in the wave-number spectra and subsequently the reflection coefficient. The aliasing results from insufficiently small sample spacing, and can be easily controlled. The truncation error results from the limited size of the measurement aperture, which may be difficult to expand for practical reasons. Therefore, extrapolation techniques based on parametric signal modeling have been developed to ''virtually'' expand the measurement aperture, thus minimizing truncation effects. However, some restrictions, particularly on spatial sampling rate, must be considered in order to obtain good estimates of the reflection coefficient. Prony series modeling is one procedure that can be used to model and extrapolate the spatial sound pressure data. Reflection coefficients estimated by using the Prony series approach will be used to illustrate this procedure.

THURSDAY MORNING, 4 DECEMBER 1997 CALIFORNIA ROOM, 8:00 A.M. TO 12:00 NOON

Session 4aUW

Underwater Acoustics: Matched-Field Processing and Time Reversal

Hee Chun Song, Chair
Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0238

Contributed Papers

8:00

4aUW1. Matched-field processing for multitonal source localization. Zoi-Heleni Michalopoulou (Ctr. for Appl. Mathematics and Statistics, New Jersey Inst. of Technol., Newark, NJ 07102)

Recent studies with simulations and real data have shown that multiple frequency information can lead to increased robustness in matched-field estimation. This observation generates the question of how this information should be combined for best results. In this work an estimator is proposed that is suitable for matched-field inversion involving multitonal sources. The processor operates on the basis of an interfrequency spatially coherent framework. Designed for passive problems, it does not require knowledge of the source spectrum, an advantage over other coherent schemes. Moreover, the processor has the additional advantage of a simple implementation. Using synthetic data, the new coherent processor was compared to the conventional Bartlett processor with incoherent averaging across frequencies. The results indicate that the new processor is, in general, preferable to the conventional Bartlett processor. Interestingly, the superiority of the coherent processor is most pronounced when the receiving array undersamples the propagation medium, making the processor attractive for its cost effectiveness. [Work supported by ONR.]

8:15

4aUW2. Range-dependent multitone matched-field processing in SWellEX-3. Phil W. Schey, Newell O. Booth, Richard T. Bachman (NCCOSC RDTE DIV, Code D881, San Diego, CA 92152-5001), and William S. Hodgkiss (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0701)

SWellEX-3 was an experimental study of shallow-water propagation and matched-field processing conducted 10 km off San Diego in July, 1994. Data were recorded on a 64-element, 118-m aperture vertical line array anchored in 200 m of water. A cw source emitting tonals from 53–197 Hz was towed over tracks where the water depth varied from 200–75 m. Examples of range-dependent matched-field processing of this multitone signal using the finite-element parabolic equation program for propagation modeling will be presented.

8:30

4aUW3. Evaluation of different techniques in matched-field processing using measured data. Ahmad T. Abawi, Newell Booth, and Phil Schey (Ocean and Atmospheric Sci., Div. NCCOSC RDTE DIV, Code D881, San Diego, CA 92152-5001)

The data from May 1996 Shallow Water evaluation cell Experiment (SWellEx-96) are used to evaluate the performance of commonly used processors in matched-field processing (MFP). The objective of SWellEx-96, which was conducted west of Point Loma in shallow water, was to investigate source tracking using matched-field processing and low-level signal detection. Data were recorded on a 64-element 118-m vertical line array and on an identical 45-deg tilted line array anchored in 200 m of water. A cw source emitting tonals from 49–388 Hz was towed over tracks ranging up to 6 km from the array. Signals were transmitted in group of frequencies. The source level in each frequency group decreased by 3 dB every 3 Hz, thus making it possible to study MFP as a function of the signal-to-noise ratio (SNR). Various figures of merits have been used to evaluate the performance of the Bartlett, the minimum variance distortionless response (MVDR) as well as MVDR with white-noise gain-constraint (MVDRWNC) processors.

8:45

4aUW4. Broadband signal detection: Comparison of vertical aperture arrays using adaptive matched-field processing and a horizontal line array using adaptive plane-wave beamforming. Newell O. Booth, Phil W. Schey (NCCOSC RDTE DIV, Code D881, San Diego, CA 92152-5001), and W. S. Hodgkiss (Scripps Inst. of Oceanogr., San Diego, CA 92152-6400)

Detection and localization of low spectral-level, broadband signals in the presence of surface interference are described and illustrated with examples obtained in the May 1996 Shallow Water evaluation cell Experiment (SWellEx-96). The experiment was carried out west of Point Loma in 200-m water of complex bathymetry. A 118-m vertical line array (VLA) was deployed next to an identical line array tilted at 45 deg (Tilted VLA). A 240-horizontal line array was also deployed nearby. The detection performance of each of the vertical arrays with adaptive broadband matched-field processing (MFP) is measured and compared with the performance of the horizontal array with adaptive broadband plane-wave beamforming.

9:00

4aUW5. Detection of low-level broadband signals using adaptive matched-field processing with horizontal arrays. Newell O. Booth, Phil W. Schey (NCCOSC RDTE DIV, Code D881, San Diego, CA 92152-5001), and W. S. Hodgkiss (Scripps Inst. of Oceanogr., San Diego, CA 92152-6400)

The detection performance of low spectral-level, broadband signals from a submerged acoustic source in the presence of surface interference is described and illustrated with examples obtained in the May 1996 Shallow Water evaluation cell Experiment (SWellEx-96). The experiment was carried out west of Point Loma in 200-m water of complex bathymetry. Adaptive broadband matched-field processing (MFP) was performed on a 240-m aperture horizontal line array. Detection and localization in range, depth, and azimuth is illustrated at all bearings except near the array broadside. The effects of horizontal aperture on MFP robustness and sidelobe levels are also discussed.

9:15

4aUW6. Broadband matched-field processing. Arthur B. Baggeroer (MIT, Cambridge, MA 02139)

Several algorithms have been proposed for broadband matched-field processing; however, the issue of signal coherence across frequency and its implications for the signal coherence has not been treated consistently. If one uses a wide sense stationary random process model for a passive source, then spectral bands are uncorrelated, and one is led to an incoherent sum of energy of the optimal beamformer across frequency bins. Active sources imply a model where there is signal coherence across frequency which can be characterized using the two-frequency coherence function. Here, it is important to preserve phase in the output of the optimum beamformer since it contains the group speed information of the several rays and/or modes of the propagation. Models for transient signals fall in between the passive and active ones. The limited duration of the transient contains multipath/multimode information; however, the signal waveform is usually not known so matched filtering is not possible. The presentation will relate the issues of signal coherence broadband MFP algorithms and provide examples of the performance under different model assumptions.

9:30

4aUW7. Matched-field processing using Bartlett sidelobe interference structures. Aaron M. Thode, W. A. Kuperman, and Gerald L. D'Spain (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0205)

Waveguide invariants [e.g., G. A. Grachev, Acoust. Phys. **39**(1), 33–35 (1993)] predict the local behavior of interference maxima/minima of acoustic intensity in the frequency-range plane. These invariants are independent of the detailed environment of the waveguide. As the Bartlett matched-field processor (MFP) ambiguity surface has a physical interpretation in terms of a time-reversed acoustic field, with the sidelobes analogous to local interference maxima, these invariant concepts are reformulated for application to MFP. The theory predicts that when a series of perfectly matched range-independent Bartlett surfaces computed over small increments of frequency are stacked and then sliced along a constant depth, the resulting sidelobe contours in the range-frequency plane display interference patterns that converge to the correct range. Indeed, these patterns are demonstrated to exist in simulated data, broadband source tows, and blue whale vocalizations recorded off San Miguel Island last year. Using the Radon transform, it is even possible to localize a source at the correct range and depth using only the trajectories of the sidelobes of the MFP surfaces. The robust nature of the range-frequency invariant suggests that incoherent postprocessing based on this concept might yield results insensitive to environmental mismatch. [Work sponsored by ONR.]

9:45

4aUW8. Coherent and incoherent broadband matched-field processing with SWellEX-96 data. Gregory J. Orris, Michael Nicholas, and John S. Perkins (Naval Res. Lab., Washington, DC 20375)

Several approaches to passive broadband matched-field processing have been proposed. In most cases, an incoherent average over a narrow band is quite successful in driving down unwanted sidelobes. The ability of coherent broadband techniques to out perform this simple approach may require additional phase information about the source that, in general, may not be available. In this paper, a fully coherent method that can out perform an incoherent average is derived, although its practical application is problematic. In other words, the simple incoherent average may be the most efficient approach. These ideas are demonstrated using data from the SWellEx-96 experiment. Given the environmental conditions under which these data were taken, the new coherent approach has some physical advantages over other broadband methods since it more efficiently uses variations of range dependence exhibited by the different frequency components of the signal. The net result is that the range, depth, and azimuth ambiguities have comparable spatial extent.

10:00–10:15 Break

10:15

4aUW9. Absolute units for acoustic pressure levels: Shades of Carl Eckart and Bob Urick [ASA medalists as Pioneers in Underwater Acoustics, 1973 and 1988, respectively]. F. H. Fisher (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0701)

In *Principles of Underwater Sound* (1946), edited and largely written by Carl Eckart, he showed what acoustic pressure levels in dynes per square centimeter (rms) meant in decibels (air and water) compared with one another and also in watts per square centimeter. In *Principles of Underwater Sound* by Bob Urick, 3rd ed. (1983), Sec. 1.6 deals very clearly with absolute units of intensity with respect to the new standard pressure reference level of 1 (μPa) = 10^{-5} dyn/cm^2 (rms) for both air and underwater acoustics; see ''Preferred Reference Quantities for Acoustic Levels,'' ANSI S1.8-1969(R1974). In physical units this intensity amounts to 0.67×10^{-22} W/cm^2. A few examples will be given of difficulties encountered by not including absolute units, most recently with respect to whales and sound sources radiating only 250 W. The inclusion of absolute units would facilitate understanding of acoustics to the scientific, technical, and public communities.

10:30

4aUW10. Near-surface source depth resolution in a shallow-water environment. W. S. Hodgkiss, J. J. Murray (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0701), N. O. Booth, and P. W. Schey (NCCOSC RDTE DIV., Code D881, San Diego, CA 92152-5001)

The depth discrimination capability of matched-field processing provides the possibility of separating surface from submerged sources of acoustic energy. Data collected in 200-m water during SWellEx-96 will be used to illustrate the achieved depth resolution in a downward refracting, shallow-water environment. During the event analyzed, two sources were towed simultaneously. The nominal depths of the deep and shallow sources were 60 and 10 m, respectively. Both projected unique, multitone transmissions with the deep source covering the 50–400 Hz band and the shallow source covering the 100–400 Hz band. In addition, the source ship herself (R/V SPROUL) had a detectable radiated acoustic signature across this same frequency region. The data were received by a 120-m aperture, 64-element vertical array deployed from the R/P FLIP. Broadband adaptive matched-field processing has been carried out on the data. The depth resolution characteristics of each of the three sources of acoustic energy will be presented. [Work supported by ONR, Code 321US.]

10:45

4aUW11. Source localization in shallow water with internal waves. Kwang Yoo and T. C. Yang (Naval Res. Lab., Washington, DC 20375)

Source localization can be a difficult problem in temporally and spatially varying shallow-water environments. The SWARM 95 data indicate that the acoustic fields at 225 and 400 Hz change drastically in the time scale of several minutes. Since it is practically impossible to model the replica field at this time scale, one expects degraded performance for matched field processing (assuming a time-independent replica field). An alternate processing scheme is proposed here. First it is shown that internal waves cause mode conversions and that mode conversions degrade matched-field source localization. This is illustrated for the Garret–Munk internal waves, the solitary internal waves and the combination of both. It is then shown that source localization can be achieved by using only the uncoupled modes. This concept was implemented using matched-beam processing. The beam filter serves to suppress the contribution of the coupled modes. It is shown that this technique significantly improves the localization performance over conventional matched-field processing. [Work supported by the Office of Naval Research.]

11:00

4aUW12. Time-reversal focusing with less than a full water column source array. W. S. Hodgkiss, W. A. Kuperman, H. C. Song (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0701), T. Akal, C. Ferla (SACLANT Undersea Res. Ctr., 19138 La Spezia, Italy), and D. R. Jackson (Univ. of Washington, Seattle, WA 98105)

Recent time-reversal mirror (TRM) experiments conducted in the Mediterranean Sea [Kuperman *et al.*, J. Acoust. Soc. Am. **101**, 3088(A) (1997)] have demonstrated the spatial focusing and temporal stability of the phase conjugation process over ranges of up to 30 km in a shallow-water waveguide. In the May 1997 experiment, the vertical source-receive array (SRA) consisted of 23 sources and receivers spanning 77 m of a 125-m water column. One series of tests involved a probe source (PS) and 94-m aperture vertical receive array (VRA) located 15 km from the SRA. Various subsets of the SRA sources were used for retransmitting the time-reversed signals received by SRA receivers from the PS. With a single source, the results essentially were the same as those reported previously [Parvulescu and Clay, Radio Elec. Eng. **29**, 223–228 (1965)]. Temporal compression is achieved but vertical spatial focusing is not. By using all 23 sources, the retransmitted field focuses distinctly at PS with levels above and below the focal region suppressed by more than 15 dB. Significant spatial focusing also was achieved with as few as six sources where these levels were suppressed by more than 10 dB. [Work supported by ONR, Code 321US.]

11:15

4aUW13. Comparisons of computed and exact time-reversing array retrofocusing in a shallow-water sound channel. Michael R. Dungan and David R. Dowling (Dept. of Mech. Eng. and Appl. Mech., Univ. of Michigan, Ann Arbor, MI 48109)

A time-reversing array (TRA) can retrofocus acoustic energy to the location of its original source even when the acoustic environment is complex and unknown. However, predictions of array retrofocusing performance require an acoustic propagation model for the environment. Results are presented from a study of the sensitivity of TRA retrofocusing to acoustic propagation modeling. Results from a modal sum propagation model and a parabolic equation (PE) code (RAM by M. D. Collins) are compared for a perfect waveguide. The influence of waveguide complexity on retrofocusing is determined by adding range dependence and realistic bottom properties to the computed PE fields. Variations in acoustic frequency and sound source depth are also explored. It is found that acoustic penetration into the bottom primarily influences sound pressure levels but does not strongly influence the size of the TRA retrofocus. In addition, small changes in the waveguide's range dependence are found to have little effect on retrofocus properties. [Work supported by the Office of Naval Research.]

11:30

4aUW14. Variable range focusing in a time-reversal mirror. Hee Chun Song, William A. Kuperman, William S. Hodgkiss (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238), T. Akal, C. Ferla (SACLANT Undersea Res. Ctr., 19138 La Spezia, Italy), and Darrell R. Jackson (Univ. of Washington, Seattle, WA 98105)

Recently, a time-reversal mirror (or phase conjugate array) was demonstrated experimentally in the ocean, which spatially and temporally re-

focused an incident acoustic field back to the original position of the probe source. [Kuperman *et al.*, accepted for publication in J. Acoust. Soc. Am.]. Here, this waveguide time-reversal mirror technique is extended to refocus at ranges other than that of the probe source. This procedure is based on the acoustic-field invariant property in the coordinates of frequency and range in an oceanic waveguide [G. A. Grachev, Acoust. Phys. **39**(1), 33–35 (1993)]. This technique has been verified in a second phase conjugate experiment conducted in the Mediterranean Sea in May 1997. The theory behind variable range focusing will be presented along with experimental results demonstrating the ability to shift the focal range ±1 km for a probe source in the vicinity of 15 km from the source–receive array.

11:45

4aUW15. Acoustic time reversal in nonlinear wave propagation. Ibrahim M. Hallaj, Steven G. Kargl (Appl. Phys. Lab, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105), and Paul E. Barbone (Boston Univ., Boston, MA 02215)

The application of acoustic time reversal in nonlinear underwater propagation is investigated theoretically and numerically. An acoustic-wave equation containing thermal and viscous absorption as well as finite-amplitude terms is presented. The behavior of such a system under time reversal is discussed. Theoretical results and numerical examples are presented, showing when and why the propagation is time reversible, and when the equation is no longer time invariant. [IMH and SGK's work is supported by the Office of Naval Research.]

THURSDAY MORNING, 4 DECEMBER 1997 SHEFFIELD ROOM 8:00 TO 9:30 A.M.

Meeting of the Standards Committee Plenary Group

S1 Acoustics U.S. Technical Advisory Group (TAG) for IEC/TC 29 Electroacoustics and ISO/TC 43 Acoustics

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 and ISO/TC 94/SC12 Hearing Protection

An initial meeting of the Standards Committee Plenary Group will precede the meetings of the Accredited Standards Committees S1, S3 and S12, to meet in the following sequence on the same day: S12—9:30 a.m. to 11:30 a.m.; S1—2:00 p.m. to 3:15 p.m.; S3—3:30 p.m. to 4:45 p.m.

(Note separate listings for these Standards Committee Meetings in the program)

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 activities including reports on recent meetings and planning for forthcoming meetings.

Members of S2 on 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 on Wednesday, 3 December 1997 at 9:00 a.m.

The U. S. Technical Advisory Group Chairs for the international Technical Committees and Subcommittees under ISO and IEC, and parallel to S1, S2, S3 and S12 are:

P. D. Schomer, U.S. Technical Advisory Group (TAG) for ISO/TC 43 Acoustics and ISO/TC 43/SC1 Noise

H. E. von Gierke, Vice Chair U.S. Technical Advisory Group (TAG) for ISO 43 Acoustics and ISO/TC 43/SC1 Noise

E. H. Berger, Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 94/SC12 Hearing Protection

J. Erdreich, U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC4 Human Exposure to Mechanical Vibration and Shock

V. Nedzelnitsky, U.S. Chair, Technical Advisory Group (TAG) for IEC/TC 29 Electroacoustics

Meeting of Accredited Standards Committee S12 on Noise

P. D. Schomer, Chair S12, and Chair U.S. Technical Advisory Group (TAG) for ISO/TC 43/SC1, Noise
U. S. CERL, P.O. Box 9005, Champaign, Illinois 61826-9005

B. M. Brooks, Vice Chair S12
Brooks Acoustics Corporation, P.O. Box 3322, Vernon, Connecticut 06066

H. E. von Gierke, Vice Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 43/SC1, Noise
1325 Meadow Lane, Yellow Springs, Ohio 45387

E. H. Berger, Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 94/SC12, Hearing Protection
E-A-R/Aearo Company, 7911 Zionsville Road, Indianapolis, Indiana 46268-1657

Accredited Standards Committee S12 on Noise. Working group chairs will report on their progress for the production of noise standards.

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 4pAA

Architectural Acoustics: Paul S. Veneklasen Memorial Session

Leo L. Beranek, Cochair
975 Memorial Drive, Suite 804, Cambridge, Massachusetts 02138-5755

David Lubman, Cochair
D. Lubman & Associates, 14301 Middletown Lane, Westminster, California 92683

Contributed Papers

1:25

4pAA1. Paul S. Veneklasen, Father, mentor. Lee H. Veneklasen (3445 Badding Rd., Castro Valley, CA 94546)

Early recollections included making things, riding bikes, sailing, boy scouts, skiing, many of the usual things that a father does with his son, but my father, Paul was different. He wanted us to learn why and how did it work; was there a better way of doing it? This atmosphere created a desire for learning and innovation. My father encouraged my education and revelled as I approached his field of electron-beam lithography from experiment and the machine shop, much as he approached acoustics. My brother, Mark, a wonderful craftsman, developed a new French Horn, with the help of my father's patient advice and occasional prodding. A few anecdotes illustrate a lifetime of memories.

1:40

4pAA2. Paul S. Veneklasen: A glimpse into his early years at UCLA, 1938–1942. Robert S. Gales (1645 Los Altos Rd., San Diego, CA 92109)

Paul joined the acoustics program of the UCLA Physics Department in 1938, shortly after receiving his M.S. in physics and mathematics from Northwestern University. He was a spirited contributor to Professor Vern Knudsen's weekly seminars dealing with nearly all aspects of acoustics: propagation, architectural, hearing, music, etc. Here, he shared ideas with other students: Dick Bolt, Cyril Harris, Bob Leonard, Walter Rosenblith, Izzy Rudnick, Ludwig Sepmeyer *et al.*, and mentors: Leo DelSasso, Vern Knudsen, and Norman Watson. In 1941 Paul became a key contributor to a new research effort to develop improved hearing protection devices as military acoustics became important in those pre-World War II years. This paper is a brief anecdotal glimpse into this short period of Paul's early professional life.

1:50

4pAA3. Paul S. Veneklasen: The War Years, 1941–1945. Leo L. Beranek (975 Memorial Dr., Ste. 805, Cambridge, MA 02138)

Paul was a senior research associate at the Electro-Acoustic Laboratory (directed by Leo L. Beranek) and the Psycho-Acoustic Laboratory (directed by S. Smith Stevens) during World War II. This paper discusses his contributions during this period to the hearing protection of military personnel and to the development of special electronic equipment for the precision measurement of the properties of sound fields.

4p THU. PM

2:00

4pAA4. Paul S. Veneklasen, 1945–1996, an overview. Jerry P. Christoff (Veneklasen Assoc., 1711 Sixteenth St., Santa Monica, CA 90404)

The author joined Paul S. Veneklasen & Associates and the Western Electro-Acoustic Laboratory in 1956, the day after his graduation from UCLA with a B.S. in Applied Physics and a specialization in acoustics. During the next forty years, he was privileged to work side by side with Paul Veneklasen on a variety of projects ranging from speech communication in noise, acoustical model testing, and a wide variety of architectual acoustical commissions. Most people think of Paul in only one field of endeavor, but those who knew him best knew that as a physicist, he would be interested and try anything to make an advance in technology. The paper provides a brief chronology of his major interests, projects, patents, and unpublished ideas.

2:20

4pAA5. Paul S. Veneklasen, contributions to electroacoustic system and design. Jose C. Ortega (Veneklasen Assoc., 1711 Sixteenth St., Santa Monica, CA 90404)

Paul Veneklasen's interest in acoustics was wide ranging. Research and development in electroacoustics occupied his endeavors in the early years. This paper will present a survey of his work. Topics to be presented include development of a series of condenser microphones and loudspeaker systems at the Altec Lansing Corporation, and development of instrumentation to measure the sound pressure level using the Western Electric 640 AA condenser microphone. Projects of a more unusual nature included the development of a Fabric Acoustimeter or ''Fuzz Feeler'' to measure by acoustic means the ''handle'' of a fabric, development of the tooth and forehead microphones for detecting and transmitting voice signals in very high-level noise enviroments and research and development of the electronic stethoscope. This instrument extended the frequency range of heart sound analysis. Paul's interest and pioneering work in auditorium acoustics led to contributions to the development of instrumentation for use in acoustical scale modeling research.

2:40

4pAA6. Paul S. Veneklasen, contributions to psychological acoustics and speech communication systems. Jerry P. Christoff and Jose C. Ortega (Veneklasen Assoc., 1711 Sixteenth St., Santa Monica, CA 90404)

Early in the authors' careers working with Paul Veneklasen at the Western Electro-Acoustic Laboratory, the authors were involved with programs dealing with speech communication in high-noise fields and the development of a new ear plug, possibly to replace the venerable V51-R ear defender that was developed by Paul Veneklasen, Vern O. Knudsen, and others during World War II. His later interests were involved with the perception of lateral reflections in rooms, which Knudsen called ''envelopment,'' primarily through research in his patented Auditorium Synthesis demonstration room. Highlights of this work are recounted including the novel forehead, ear, and tooth microphones.

2:55

4pAA7. Paul S. Veneklasen, contributions to enviromental noise control. Jose C. Ortega (Veneklasen Assoc., 1711 Sixteenth St., Santa Monica, CA 90404)

Paul had a deep and continuing interest in controlling ennviromental noise. He was instrumental in organizing one of the earliest symposiums on urban noise. His ''City Noise—Los Angeles'' paper in *Noise Control Magazine* in July 1956 is filled with prophetic comments on urban noise and its potential effect on people. Paul was a guiding influence on the development of a Model City Noise Ordinance commissioned by the Department of Health of the State of California. Paul not only reported on enviromental noise but was actively engaged in solving noise problems. His work on the development of a ground run-up suppressor system design for the Northrop F-89 aircraft is a text book example of defining, understanding, and the developing methods to reduce enviromental noise from a very high-level noise source.

3:10–3:25 Break

3:25

4pAA8. Acoustical modeling and simulation in the 1960s: Paul Veneklasen's quest for ''envelopment'' and correlations with auditorium design. Jerald R. Hyde (Jerald R. Hyde Acoust., Box 55, St. Helena, CA 94574)

Paul's favorite ''in-house'' project in the 1960's was his auditorium synthesis system. He devised an electro-acoustic laboratory listening environment which simply divided a test sound field, excited by an anechoic recording, into the three basic components called ''direct,'' ''envelopment,'' and ''reverberance.'' The envelopment factor was the summation of early delayed reflections which surrounded the listener at the sides via loudspeakers. Paul was primarily looking for how much envelopment, or lateral early energy, a listener preferred relative to the direct sound level. At around the same time, he was using a nitrogen jet ultrasonic noise source to model test exactly how much lateral early energy was possible. He found that in a conventional hall with the best possible lateral energy design, the amount of envelopmental energy available at most seats was around 3 to 5 dB lower than ''desired.'' Concurrently, Marshall and then Barron extended the previous work of Keet and Kuhl in developing the parameter called ''lateral energy fraction,'' which essentially quantified Veneklasen's envelopmental energy. Computer systems later allowed the actual measurement of typical lateral energy values in concert halls. Veneklasen's findings will be compared to the actual measured values for hall's of well-known quality.

3:40

4pAA9. Paul Veneklasen and the design and engineering of the dampened mild carbon steel stage acoustic shell: A reminicence. George C. Izenour (16 Flying Point Rd., Stony Creek, CT 06405)

Paul and I were introduced to each other in the early 1960s by Vern Knudsen with whom I was then collaborating on Frank Lloyd Wright's last public building, The Grady Gammage Auditorium on the Campus of Arizona State University (1964). My initial collaboration with Paul, with Vern Knudsen looking over both our shoulders, was the multipurpose arena on the campus of Washington State University. Simultaneously with the Washington State Project (1973) there were two other collaborations: (1) the remodeling-restoration of the Eastman theater on the campus of the University of Rochester (1972) and (2) the Hancher auditorium on the campus of the University of Iowa (1973). All told, my three collaborations, with Paul as acoustics consultant and myself as theater consultant, extended over a period of about 5 years (1968–1973). It is our collaboration on the Eastman project that is the subject of this paper.

4:00

4pAA10. Paul S. Veneklasen, contributions to architectural acoustical design, 1960–1996. Jerry P. Christoff and James A. Good (Veneklasen Assoc., 1711 Sixteenth St., Santa Monica, CA 90404)

Paul Veneklasen will be remembered most for his contributions to architectural acoustics, especially completed, highly acclaimed auditoriums, ranging from the Seattle Opera House and Los Angeles Music Center to Wolf Trap Farm Park to name a few. But he also left a legacy of hundreds of radio, television, and recording studios, and virtually every other building type from office buildings to convention centers and even jet engine test cells and noise suppressors. But less known are his development of orchestra enclosures, pit lifts, multilayer gypsum board construction, damped glass panels, floating floors, and sound doors incorpo-

rating magnetic seals and acoustical filters. His research extended from optimizing sight lines and seating arrangements in auditoriums, acoustical modeling, and theater chairs. Highlights of Paul Veneklasen's well-known projects, developments, and research are presented.

4:20

4pAA11. The auditorium measurement chain: A study of the use of auditorium synthesis in the design of concert halls. John J. LoVerde (Veneklasen Assoc., 1711 Sixteenth St., Santa Monica, CA 90404)

In the early 1960s, Paul Veneklasen began to develop an idea to synthesize an indoor acoustical space coming to fruition with the development and construction of an electroacoustic system called Auditorium Synthesis. This system allows the listener to preview various types of spaces prior to construction [Paul S. Veneklasen, Int. Symposium on Arch. Acoustics, 21–43 (1974)]. Listener preferential settings including source strength (G), early decay time (EDT), early/late energy (C), and lateral energy fraction from the Otis A. Singletary Center for the Arts (Lexington, Kentucky). A comparison of preferred orchestral settings and modeled seat locations will be performed. With impulse response data from the completed concert hall at the Singletary Center for the Arts, a comparison will be shown of newer parameters between the modeled seat locations from Auditorium Synthesis and concert hall data.

4:35

4pAA12. The acoustical modeling of Paul Veneklasen. Gary W. Siebein and Martin A. Gold (Univ. of Florida, Gainesville, FL 32611)

The pioneering acoustical modeling work of Paul Veneklasen has served as an inspiration to much current work in the area, including that conducted at the University of Florida. Paul constructed small models of spaces he was designing and conducted a variety of acoustical tests in them to assist in making design decisions on many innovative rooms. He determined criteria by which to evaluate the acoustic response of the rooms obtained from the models from listening to simulated sounds in his laboratory. These criteria included: providing a strong direct sound throughout the room; providing reflections from overhead within 40 ms or less of the direct sound for clarity and loudness; providing lateral reflections within 80 ms of the direct sound for envelopment; and providing a rich reverberant sound across a full range of frequencies. These criteria are still used today as the basis for interpreting the impulse responses obtained in the design process of rooms. The experiments he engaged in as a part of the process of many rooms involved research and design components including listening, modeling, designing, and building can be traced in concept to the work of Wallace Sabine. This process has become an established approach to acoustical design today.

4:50

4pAA13. Paul S. Veneklasen, mentor and associate. Jerry P. Christoff and Jose C. Ortega (Veneklasen Assoc., 1711 Sixteenth St., Santa Monica, CA 90404)

The authors both arrived at Paul S. Veneklasen & Associates and Western Electro-Acoustic Laboratory in 1956 and 1957, fresh from UCLA's acoustics program which was part of applied physics. In an era when most people change jobs numerous times during their careers, what was the glue that kept the authors there? Security? Technical challenge? Opportunity to work with a truly creative, renaissance person? Probably all of the above. But more than acoustics and business were learned from Paul. A few anecdotes illustrate the affection for Paul and the legacy he left.

THURSDAY AFTERNOON, 4 DECEMBER 1997 WINDSOR ROOM, 1:00 TO 2:45 P.M.

Session 4pABa

Animal Bioacoustics: Reptile and Amphibian Bioacoustics

Peter M. Narins, Chair
Department of Physiological Science, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, California 90095

Chair's Introduction—1:00

Invited Papers

1:05

4pABa1. Temporal processing in water and air: Developmental changes in synchronized response to AM signals across metamorphosis in the bullfrog. Seth S. Boatright-Horowitz and Andrea M. Simmons (Dept. Neurosci., Brown Univ., Box 1953, Providence, RI 02912)

Acoustic limitations on propagation of low-frequency complex communication sounds make the amplitude modulation (AM) rate an important auditory feature for both aquatic larval (tadpole) and partly terrestrial (adult) bullfrogs. Phase-locked responses to AM signals were recorded from neurons in the auditory midbrain during metamorphic and postmetamorphic development. The best modulation frequency (BMF) and maximal significant AM rate are significantly negatively correlated with the developmental stage. Pre- and early prometamorphic tadpoles show phase locking to higher AM rates than postmetamorphic animals. Late prometamorphic (deaf period) tadpoles show a substantial reduction in auditory sensitivity, lower BMFs, reduced bandwidths of modulation transfer functions (MTF), and loss of significant phase locking. By the onset of the metamorphic climax, there is a recovery of auditory sensitivity and phase-locked responding, with late climax tadpoles showing auditory sensitivity similar to that of postmetamorphic

frogs. The changes in neural responding across metamorphosis are correlated with morphological changes in the auditory periphery and central auditory nuclei, and match differences in the acoustic environment as the animal shifts from an aquatic to a more terrestrial lifestyle. [Research supported in part by an NIH research Grant NS28565 (AMS) and an NSF Graduate Fellowship (SBH).]

1:25

4pABa2. Sound radiation and postglottal filtering in frogs. Alejandro P. Purgue (Dept. Physiol. Sci., UCLA, Los Angeles, CA 90095, apurgue@ucla.edu)

This work presents a comparison across selected species of several aspects of the mechanism of sound radiation in frogs. Measurements of the magnitude of the transfer function of the radiating structures show that the structures radiating the bulk of the energy present in the call vary depending on the species. Bullfrogs (*Rana catesbeiana*) radiate most of the energy (89% sound level) in their calls through their eardrums. Additionally, the transfer function of the eardrums displays peaks coincident with those observed in the mating and release call of this species. The vocal sac and gular area contribute energy only in the lower band (150–400 Hz). The ears are responsible for radiating additional frequency bands to the ones radiated through the vocal sacs. In *Rana pipien*s, the ears also broadcast a significant portion of the energy present in the call (63% sound level), but the frequencies of the aural emissions are a subset of those frequencies radiated through the vocal sacs. Finally, the barking treefrog (*Hyla gratiosa*) appears to use two different structures to radiate his call. The low-frequency band is preferentially radiated through the lungs while the high-frequency components are radiated through the vocal sac.

1:45

4pABa3. Acoustic tuning mechanisms in reptilian hair cells. Ruth Anne Eatock (The Bobby R. Alford Dept. of Otorhinolaryngology and Communicative Sci., Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, eatock@bcm.tmc.edu)

Tuning by hair cells can be mechanical or electrical. The best-studied examples of each class are from reptilian ears. In the free-standing region of the alligator lizard's cochlea, where hair cells have acoustic characteristic frequencies (CF_a's) that range systematically from 1 to 4 kHz, acoustic tuning can be explained by micromechanical resonances of the hair bundles [D. M. Freeman and T. F. Weiss, Hear. Res. **48**, 37–68 (1990)]. In the turtle cochlea, hair cells have CF_a's from 30 to 700 Hz and the acoustic tuning derives from electrical resonance of the hair cell membrane, reflecting the interplay of capacitive current and several ionic currents [Y.-C. Wu *et al.*, Prog. Biophys. Mol. Biol. **63**, 131–158 (1995)]. Tonotopic variation in CF_a arises from variation in hair bundle height in the lizard and from variation in ion channel kinetics and number in the turtle. Coupled mechanical and electrical tuning [T. F. Weiss, Hear. Res. **7**, 353–360 (1982)] seems likely but has not been demonstrated. Hair bundle height varies with Cf_a in the turtle cochlea, suggesting a micromechanical contribution. Hair cells dissociated from the liquid cochlea resonate electrically, but at frequencies tenfold lower than their CF_a's [R. A. Eatock *et al.*, J. Neurosci. **13**, 1767–1783 (1993)].

2:05

4pABa4. Automated bird songs recognition using dynamic time warping and hidden Markov models. Joseph A. Kogan and Daniel Margoliash (Dept. of Organismal Biol. and Anat., 1027 East 57th St., The Univ. of Chicago, Chicago, IL 60637, joseph@modeln.uchicago.edu)

The performance of two well-known recognition techniques in speech adapted to the automated recognition of bird song units from continuous recordings is studied. The advantages and limitations of dynamic time warping (DTW) and hidden Markov models (HMMs) are evaluated on a large database of songs of two species: zebra finches (*Taeniopygia guttata*) and indigo buntings (*Passerina cyanea*), which have different types of vocalizations and have been recorded under different laboratory conditions. The recognition performance of these methods is also assessed with regard to song signal representation (including representations commonly used in speech recognition), model structure, and sensitivity. Depending on the quality of recordings and song complexity, the DTW-based technique gives from excellent to satisfactory results. Under more challenging conditions, such as noisy recordings or the presence in bird vocalizations of confusing transient calls, HMMs can significantly outperform DTW but requires more training examples. Even though the HMMs perform usually quite well, they often misclassify short transient calls and song units with more variable structure. To address these and other weaknesses of the studied techniques, new approaches to analyze transient and variable bird vocalizations are discussed. [Work supported in part by U.S. Army Research Office and NIH.]

2:25

4pABa5. Desert tortoises (*Gopherus agassizii*) lack an acoustic startle response: Implications for studies of noise effects. Ann E. Bowles and Scott A. Eckert (Hubbs-Sea World Res. Inst., 2595 Ingraham St., San Diego, CA 92109)

The startle reflex is little studied in nonmammalian vertebrates, but in mammals it is often characterized by increase in heart rate. Orienting can also occur, characterized by a brief decrease in heart rate. To determine whether these responses occur in a testudinate, the desert tortoise, heart rate and behavior were measured during exposures to simulated jet overflights (94.6–114.2 dB CSEL) and sonic booms (6–10.5 psf). The best sensitivities of the 14 subjects ranged from 23 dB–50 dB SPL (average 34 dB SPL at 250 Hz) measured using ABR. Initial exposure to simulated jet overflights produced a typical reptilian defensive response: freezing. Tortoises became totally immobile for periods of up to 113 min. The average heart rate showed a gradual decline (7–8%), recovering within 2–4 h. Tortoises oriented when exposed to simulated sonic booms and after repeated exposure to subsonic aircraft noise; orienting produced no detectable change in heart rate. These results suggest that (1) high-intensity transients affect desert tortoises by altering activity patterns, but do not have a direct effect on heart rate and (2) tortoises experience a physiological response that produces protracted freezing. [Work supported by USAF, Edwards Air Force Base, and F-22 SPO.]

Session 4pABb

Animal Bioacoustics: General Topics: Effects of Noise on Animals

Ann E. Bowles, Chair
Hubbs-Sea World Research Institute, 2595 Ingraham Street, San Diego, California 92109

Contributed Papers

2:50

4pABb1. Effects of fixed-wing military aircraft noise on California gnatcatcher reproduction. Frank T. Awbrey and Don Hunsaker II (Hubbs Sea World Res. Inst., 2595 Ingraham St., San Diego, CA 92109)

To test the assumption that high levels of aircraft noise impede bird reproduction, noise analyzers were placed for 1 week in the nesting territory of each of 39 California gnatcatcher pairs on Naval Air Station Miramar. The 1-week average sound levels (7DL) recorded in those nesting territories were then related to the number of nest attempts; number of eggs laid; number of chicks hatched; number of chicks fledged; and number of eggs, chicks, and fledglings per nest attempt. Nest attempts and eggs laid have weak negative correlations ($p=0.14$ and 0.28) with 7DL. That is, the birds may tend to build fewer nests and lay fewer eggs in noisier areas, which is consistent with the common observation that bird nesting is more easily disturbed before eggs are laid than after. None of the other indicators is correlated with sound levels. Once a nest is established, with eggs in it, military aircraft noise has no detectable influence on reproductive performance. Gnatcatchers reproduced in places where 1 HL exceeds 80 dB for several hours every day. If fixed-wing aircraft noise impedes California gnatcatcher reproduction, it is overwhelmed by such factors as disturbance, predation, weather, edge effects, and differences in quality of habitat.

3:05

4pABb2. Response of elephant seals to acoustic thermometry of ocean climate sound transmissions. Daniel Costa, Daniel Crocker, James Gedamke, Paul Webb, Burney Le Boeuf, Danielle Waples, Sean Hayes, and James Ganong (Dept. of Biol., Univ. of California, Santa Cruz, CA 95064)

The hypothesis that northern elephant seals would respond to acoustic transmissions from the ATOC sound source was tested. Elephant seals were chosen because they have the best low-frequency hearing of any pinniped, are abundant, naturally migrate past the Pioneer Seamount, and are deep divers. ARGOS satellite tags provided information on animal location while at sea, while archival tags provided information on swim speed, time, depth, ambient acoustic environment, and ambient sound pressure levels. Instruments were deployed on 14 naturally migrating adult male elephant seals that were expected to swim near the source site, and upon 29 juvenile animals that were translocated and released 1 h prior to transmission. Measured mean intensity of ATOC exposure ranged from 120 to 135 dB for 60–90 Hz ($n=6$) compared to ambient levels of 100–107 dB (60–90 Hz). Animals did not alter return track, diving pattern, or swim speed, did not go to the surface, and often continued to dive closer to the sound source if on the descending segment of a dive. [Work funded by ONR and ARPA.]

3:20

4pABb3. Acoustic interaction of humpback whales and whale-watching boats off Maui, Hawaii. Whitlow W. L. Au (Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734) and Marsha Green (Albright College, Reading, PA 19612)

Whale-watching activities during the humpback whale annual winter sojourn in Hawaii have increased steadily over the past years. A concern of many is the possible detrimental effects of whale-watching boats on humpback whales; including the possibility of damage to their auditory system. To address this concern, the noise generated by five different types of whale-watching boats was measured, two with outboard engines, two with inboard diesel engines, and a seagoing vessel, in the near shore waters of west Maui. The first series of measurements was conducted at the peak of the whale season, and the boat noises were partially masked by humpback whales' chorusing (boat noises and whale chorus levels were on the same order of magnitude). A 1/3 octave band spectral analysis indicated whale chorus levels as high as 136 dB re: 1 μPa with no whales in sight. Throughout the day humpback whale chorusing noise was present at similar levels for miles along the same coastline. Measurements taken close to the end of the season had about a 15-dB decrease in the humpback whale chorusing levels. Considering the high level of chorusing by humpback whales, the acoustic effects of the whale-watching boats should be relatively small or negligible.

3:35

4pABb4. Synthesis of distant signatures of underwater explosions for sea mammal hearing studies. Joseph A. Clark and Jane A. Young (CDNSWC, Bethesda, MD 20084 and COMB, Ste. 236, Columbus Ctr., 701 East Pratt St., Baltimore, MD 21202)

Experiments which measure temporary threshold shifts (TTS) in marine mammals produced by loud noises provide one means for determining the effects that underwater explosions might have on sea life in the neighborhood of the explosion [S. Ridgway *et al.*, J. Acoust. Soc. Am. **101**, 3136(A) (1997)]. In order to obtain meaningful results which do not presuppose an understanding of the possible hearing damage mechanisms, the waveforms of the noises used as stimuli in the TTS measurements should reproduce the complexities of the actual sound field experienced by sea mammals at the outer edges of the region within which significant acoustic effects on hearing are expected. Multipath arrivals and significant refractive effects complicate the far-field signals produced by explosive shock tests. For example, a large number of superimposed short impulses are often observed. In this talk a system for synthesizing distant signatures of underwater explosions will be described. The computer-controlled system takes calculated pressure-time waveforms as input and generates an acoustic simulation of the input signal. Preliminary tests to validate the operation of the system will also be reported.

Session 4pABc

Animal Bioacoustics: Animal Bioacoustics Poster Session

Ann E. Bowles, Chair
Hubbs Sea World Research Institute, 2595 Ingraham Street, San Diego, California 92109

All posters will be on display from 4:00 p.m. to 6:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 4:00 p.m. to 5:00 p.m. and contributors of even-numbered papers will be at their posters from 5:00 p.m. to 6:00 p.m.

Contributed Papers

4pABc1. Bat jitter discrimination, adaptive channel equalization, and the critical interval. Richard A. Altes (Chirp Corp., 8248 Sugarman Dr., La Jolla, CA 92037, altes@msn.com) and James A. Simmons (Brown Univ., Providence, RI 02912)

When simulated echoes have a bandwidth between 20 and 80 kHz, bats (*Eptesicus fuscus*) can discriminate between pulse-to-pulse delay jitter and constant echo delay with a sensitivity of 10 ns [Simmons *et al.*, J. Comp. Physiol. A **167**, 589–616 (1990)]. To investigate this amazing acuity, high-pass and low-pass filters were inserted into the channel between the simulated target and the bat. The effect of such filters on jitter acuity was measured. The data are explained by a receiver model that attempts to equalize the channel filter. The equalizer uses a smeared version of the actual channel transfer function, where spectral smearing induces relatively slow cutoff. Spectral smearing occurs in spectrogram (short-time Fourier transform) analysis. Smearing varies inversely with the impulse response duration of the analysis filters. This impulse response duration corresponds to the critical interval, which is frequency dependent for proportional bandwidth filtering. Bat jitter discrimination data imply channel equalization with a smeared version of the actual channel transfer function. This smearing is commensurate with critical intervals that are obtained from other experiments [Simmons *et al.*, J. Acoust. Soc. Am. **86**, 1318–1332 (1989)]. Bats apparently use time-frequency echo representations to perform environmentally adaptive filtering.

4pABc2. Quantitative and qualitative analysis of the ascendent frequency whistles of tucuxi, *Sotalia fluviatilis*, in the Sepetiba Bay, Rio de Janeiro, Brazil. Sheila M. Simao, Luciana D. Figueiredo, and Salvatore Siciliano (UFRRJ, Dept. Ciencias Ambientais, Seropedica, Rio de Janeiro, 23851-970, Brazil)

Sotalia fluviatilis was the only cetacean species observed in all 16 boat surveys (82 h) conducted in the Sepetiba Bay (22° 58′, 44° 02′W), between December 1993 and August 1995. During the boat survey of 5 January, 1995—a very calm day (Beaufort = 1)—it was possible to make good subaquatic sound recordings (30 min) of a group of *ca*. 20 young and subadult dolphins. As there is little bioacoustic information on tucuxi, sonogram analysis was used to characterize the tucuxi's whistles of ascendent frequency. The qualitative analysis resulted in 866 whistles divided in 48 types. Some of these sonograms shapes match the ones presented by other cetacean species sonograms (*Tursiops truncatus, Delphinus delphis, and Globicephala macrorhynchus*). The quantitative analysis was based on: number of harmonics; number of inflexions; beginning, final, maximum, and minimum frequecies; duration; and frequency modulation. Minimum, maximum, and average values, standard deviation, and coefficient of variation were calculated for each type of whistle. The *S. fluviatilis* whistles are extremely diverse (high coefficient of variation values for each type) and frequent (38.5 whistles/min), seeming to have an important role in social relations among these animals. Such whistle diversity and quantities could be explained by the composition of the dolphin group, i.e, it was exclusively formed by young and subadult animals.

4pABc3. Quantitative and qualitative analysis of the descent, up–down, and down–up frequency whistles of tucuxi, *Sotalia fluviatilis*, in the Sepetiba Bay, Rio de Janeiro, Brazil. Tereza C. C. L. Pereira, Sheila M. Simao, and Salvatore Siciliano (UFRRJ, Dept. Ciencias Ambientais, Seropedica, Rio de Janeiro, 23851-970, Brazil)

Sotalia fluviatilis was the only cetacean species observed in all 16 boat surveys (82 h) conducted in the Sepetiba Bay (22° 58′S, 44° 02′W), between December 1993 and August 1995. Previously, *S. fluviatilis* whistles were bioacoustically poorly characterized. The objective of this research was a quali–quantitative characterization of tucuxi's acoustic emissions of descent, up–down, and down–up frequency. The sound analysis was based on sonograms from 30 min of a subaquatic recording obtained on 5 January, 1995. The qualitative analysis resulted in 390 whistles classified in 3 subclasses (descent, up–down, and down–up frequency), 21 types, and 46 subtypes. The quantitative analysis was based on: number of harmonics; number of inflexions; beginning, final, maximum, and minimum frequencies; duration; and frequency modulation. Minimum, maximum, and average values, standard deviation, and coefficient of variation were calculated for each type and subtype of whistle. The *S. fluviatilis* whistles are extremely diverse (high coefficient of variation values for each subclass, type, and subtype) and frequent (38.5 whistles/min), seeming to have an important role in social relations among these animals. Such whistle diversity and quantities could be explained by the composition of the dolphin group, i.e., it was exclusively formed by young and subadult animals.

4pABc4. Acoustic geographic variation in two populations of Delphinus delphis. Caitlyn Toropova (1264 Poplar Dr., Arcata, CA 95521)

Vocal patterns from two populations of common dolphin (*Delphinus delphis*) were compared. Field observations took place in the Channel Islands, CA and La Paz, Mexico. Whistles were recorded with a Sony HTI hydrophone (range=10–32 000 kHz) and schools were observed for group number and distance to the hydrophone by onboard observers. Sounds and spectrograms were analyzed for mean intensities in the (a) 15- to 20-kHz bandwidth, and (b) the ratio of the intensities from the 15- to 20-kHz band compared to the whole spectrum, 0–22 kHz. Significantly different ($p<0.001$) sound intensity levels were found between the La Paz and Channels Islands populations. When the 15- to 20-kHz band was

examined, the values were 104.70±1.73 dB (mean ±S.E.) and 86.80 ±0.90 dB for the La Paz and Channel Islands, respectively. When the 15- to 20-kHz band was looked at as a function of the total energy in the 0 to 22-kHz band, significantly different energy levels were present as well (La Paz=76.2%, Channel Islands=67.9%). Therefore, the two populations were shown to be significantly different (p=0.001). This suggests that acoustic geographic variation is a useful tool for distinguishing populations within a species, and raises the possibility of external influences determining vocal patterns.

4pABc5. Vocalization as an indicator for disorders in mammals. Tobias Riede, Guenter Tembrock (Inst. fuer Biol., Abt. Verhaltensphysiologie, Humboldt-Univ., Invalidenstrasse 43, 10115 Berlin, Germany), Hanspeter Herzel (Humboldt-Univ., Berlin, Germany), and Leo Brunnberg (Freie Univ., Berlin, Germany)

As in humans, diseases in animals may lead to characteristic changes of the voice. In particular, multiparametric frequency, time analysis, as well as the amount of nonlinear phenomena (frequency jumps, subharmonics, biphonation, chaos) provide information about the well-being of an animal and the progress of a disease. Vocalization of domestic animals (dogs and cats) in a university hospital and of a macaque species in a zoological garden have been recorded. First results show that diseases lead typically to shortened call length and decreasing frequency modulation in the cat. Using narrow-band spectrogram subharmonics, biphonation, frequency jumps, and deterministic chaos are identified during the recuperation process of a cat with craniocerebellar trauma. The same nonlinear phenomena appear in several dogs with dysphonia following exhaustion of the voice and in a Japanese macaque infant with an unidentified systemic disorder. The ratio of nonlinear phenomena during systemic disorders of an animal shows a species-specific pattern. There was an increase mostly of frequency jumps in the macaque infant, of frequency jumps and chaos in the dog, and of chaos and subharmonics in the cat vocalization. [Work supported by NaFoeG, Humboldt-University Berlin.]

4pABc6. Variations in the vocal repertoire of the manatee in the Caribbean (*Trichechus manatus manatus*). Jose A. Alicca, James Harvey (Moss Landing Marine Labs., P.O. Box 450, Moss Landing, CA 95039-0450), and Antonio Mignucci-Giannoni (Caribbean Stranding Network, San Juan, Puerto Rico 00937)

While different life history parameters for the manatee have been well studied in Florida, little is known about these aspects of manatee biology in the Caribbean. As part of a multifaceted study in Antillean manatee life history, the geographical and individual variability in vocal repertoires of Antillean manatees from Puerto Rico and the Dominican Republic were examined and compared to the repertoire of the Florida manatee. Differences in structural sound characteristics (frequency, duration, and relative power of each harmonic), usage (number of call types), and the associated behavior were studied from recordings of wild and captive animals. Over 100 h of recording were obtained from six different natural localities in Puerto Rico and one natural locality in Florida. Captive manatees were recorded in the Dominican Republic (1), in Puerto Rico (2), and Florida (13), in addition to recording seven semicaptive individuals in Florida. The recording of underwater vocalizations has been suggested as a useful tool for the study of marine mammals distribution. In the turbid waters of rivers, bays, and lagoons where manatees congregate in the Caribbean, these acoustic data might be useful for estimating population distribution and habitat use.

4pABc7. Is the echolocation ability of bottlenose dolphins too good? William A. Friedl (Natl. Defense Ctr. of Excellence for Res. in Ocean Sci., 73-4460 20 Queen Kaahumanu Hwy., Ste. 111, Kailua-Kona, HI 96740, billf@ceros.org)

Odontocete echolocation studies reveal a system with both great plasticity and precision, but knowledge of the animal's day-to-day use of the process is limited. High-speed cine photographs of a trained bottlenose dolphin swimming underwater show that the dolphin's head and rostrum move vertically even when the dolphin swims at modest speed. Does this movement limit or otherwise affect the echolocation precision of freely-swimming dolphins? Results from kinematic studies, combined with echolocation system descriptions from acoustic research, were examined to show if head movements might limit acoustic abilities of swimming dolphins. So far, these speculative approaches are inconclusive. The general synthesis, using disparate data from acoustic and kinematic studies, indicates that head motion associated with modest swimming speed does not degrade echolocation ability. Data from coordinated studies that involve both echolocation and movement at realistic, natural speeds are lacking. Given the difficulty involved in working with naturally swimming dolphins, this data set is likely to remain unfilled, so further conclusions will also remain speculative.

4pABc8. A source–filter model of humpback whale (*Megaptera novaeangliae*) vocal sound production. Eduardo Mercado III (Kewalo Basin Marine Mammal Lab., Univ. of Hawaii, 1129 Ala Moana Blvd., Honolulu, HI 96814, mercado@hawaii.edu)

Acoustic waveforms produced by vocalizing humpback whales are well described by a source–filter model. This model characterizes vocalizations as the output of a time-varying signal generator (the source) passed through a time-varying resonator (the filter). For most humpback whale vocalizations, the source can be modeled as either a quasiperiodic impulse train, white noise, or a combination of these two signal types; the resonator can be modeled as an all-pole filter. The source–filter model is useful because it (a) provides a way to quantitatively characterize humpback whale vocalizations that can be automated and that allows for the ''resynthesis'' of sounds; (b) facilitates the application of advanced speech processing technologies to the analysis of humpback vocalizations; and (c) can potentially provide clues about the actual mechanisms humpback whales use to produce sounds. Currently, it is not known how humpback whales internally produce sounds, although structures in the larynx are thought to be involved. Source–filter-based analysis of humpback vocalizations revealed characteristic regions of frequency enhancement (formants) that may correspond to resonant frequencies of a physiological ''filter.'' Distribution of formants varied with pitch in a manner that appeared similar to pitch-related formant variation in human singers, suggesting that both species use analogous production mechanisms.

4pABc9. A sonobouy array for two-dimensional location of dolphin vocalizations. Eric S. Howarth and R. H. Defran (Cetacean Behavior Lab., Dept. of Psych., San Diego State Univ., 5500 Campanile Dr., San Diego, CA 92182)

A four-sonobouy array was deployed 0.25 km offshore in southern California waters. The array was developed and deployed to passively locate coastal bottlenosed dolphins (*Tursiops truncatus*). Sonobouy mooring sites were selected to overlap areas shown by the Cetacean Behavior Laboratory at San Diego State University to be highly utilized by these coastal dolphins. Arrival-time-difference measurements were used to provide two-dimensional locations of underwater biological sounds such as dolphin vocalizations. Sonobouys were separated by a range of between 367 and 1585 m and the accuracy of source location depended on the relative proximity to the nearest sonobouy. Global Positioning Satellite and theodolite data were used to verify sonobouy locations and, in turn, the acoustic arrival-time differences of on-board calibration noises. Bottlenosed dolphin whistles, including mimics of the calibration sound, were among the biologically identifiable sounds tracked by the array. Other

acoustic signals recorded but not located were snapping shrimp (*Californiensis dentripes*) and sources judged to be fish. Recorded sound levels varied over a considerable range, even on identical hydrophones, and were probably due to the signal transmission distance, shallow-water propagation effects, and diurnal fluctuations in the sources.

4pABc10. Selection of instrumentation for monitoring undersea animal sounds. Daniel R. Raichel (Dept. of Mech. Eng., City College of New York, Steinman Hall, New York, NY 10031)

In order to properly select instrumentation for bioacoustic studies, a list of measurable criteria must be drawn up, but even that may not suffice unless some genuine thought is given to the possibility of mapping additional criteria, such as frequencies beyond the range usually considered. In the planning and design of the experiment, some of the criteria include: frequency and dynamic ranges, the number of channels desired (not necessarily always available!), compatibility of components (as exemplified by the matching of acoustical and electrical impedances), setup of sensing/receiving transducer arrays, mode of recording, statistical methodologies, usable computer prowess, etc. Environment constitutes an extremely important factor from two aspects: (1) It should not be adversely affected by the use and operation of the equipment; and (2) the equipment should be able to withstand rigors of temperature, humidity, vibration, and shock. Provision should be made for instrument calibration at proper intervals to ensure the integrity of obtained data. Cost is another factor that can determine the reliability of the equipment, quality of performance, ruggedness, and ease of operation.

THURSDAY AFTERNOON, 4 DECEMBER 1997 FORUM ROOM, 2:45 TO 4:00 P.M.

Session 4pEA

Engineering Acoustics: Flow Noise, Aeroacoustic Noise

Sung-Hwan Ko, Chair
Naval Undersea Warfare Center, Code 2133, Building 1171/3F, Newport, Rhode Island 02841

Contributed Papers

2:45

4pEA1. Measurement and empirical prediction of bottom surface acoustic pressures on a hover aircraft model. Herbert L. Kuntz, Floyd O. Hickmon III, and Edward P. Feltz (Lockheed Martin Skunk Works, 1011 Lockheed Way, Palmdale, CA 93599-2522, hkuntz@lmco.com)

Acoustic measurements were made on the bottom surface of the 86% scale large-scale propulsion model (LSPM) of an advanced short takeoff and vertical landing (ASTOVL) airplane. The measurements were obtained during fixed hover heights where the lower wing surface varied from 5 ft (airplane's wheels on the ground) through 36 ft above the ground for a wide range of propulsion thrust mixes and airplane control surface configurations. The complete propulsion system provided four exhausts through an integrated unit consisting of an aft located main jet engine, which powered a forward located shaft driven lift fan and was completed with two main engine bleed high-pressure bypass roll jets (in the wings). Flush surface measurements were made at 12 locations on the bottom surfaces and time correlated with the operating propulsion parameters. An empirical acoustic model was made through analyses of the geometry and the data, and was used to predict contours on the bottom of the aircraft. [This work was performed by Lockheed Martin Skunk Works at the NASA-Ames Research Center.]

3:00

4pEA2. Approximate model for sound generation due to unsteady flows in pipes. Michael Krane, Daniel Sinder, and James Flanagan (CAIP Ctr., Rutgers Univ., Piscataway, NJ 08855)

An approximate model for the generation of sound due to unsteady vortical flow in pipes of varying cross-sectional area is described. This model, which is being used in speech synthesis research, is based upon Howe's acoustic analogy. This formulation allows the sound generation to be specified in terms of properties of the unsteady behavior of a vortical flow and the potential flow solution for the pipe that would exist in the absence of vortical inhomogeneities. An approximate model for vorticity formation and evolution is presented, and the potential flow solution is obtained from the axial duct shape. Computed estimates of the time evolution of the sound generated by a confined jet passing through a pipe constriction are presented. These results are compared to experimental data. [Research supported by NSF/ARPA IRI-9314946 and ARPA DAST 63-93-C-0064.]

3:15

4pEA3. Experimental study of cavity resonance suppression methods. J. Scott Johnson and Luc G. Mongeau (Dept. of Mech. Eng., Purdue Univ., West Lafayette, IN 47907)

Flow-excited cavity resonance is a common, generally undesirable, sound-generating phenomenon that occurs in many engineering applications. In this study, low-speed cavity resonance suppression techniques were investigated. The focus was on two methods: leading edge spoilers and leading edge air mass injection. Experiments were conducted in a closed section wind tunnel, which included a secondary blowing apparatus for controlling boundary layer thickness. The cavity, attached to the wind tunnel test section, acted as a side-branch resonator. Smoke flow visualization was performed for various operating conditions. For example, a nondimensional spoiler length of 0.156 (based on cavity length) inclined at a 50-deg angle of attack typically provided a 12 dB (re: 20 μPa) reduction of the 140-dB, 124-Hz tonal cavity pulsation. Air mass injection through a perforated array at a rate of 2.8 liter/s yielded a 27-dB reduction. In general, the results suggested that such leading edge suppression techniques yield greater orifice shear layer thicknesses, similar to that of increasing the boundary layer thickness. The thickened shear layer may cause greater vortex diffusion, a key parameter establishing the level of the cavity pressure fluctuations. Diffuse vortices lead to weaker cavity pressure oscillations and lower vortex convection velocities.

3:30

4pEA4. Turbulent flow noise estimate by use of a velocity sensor embedded in a three-layered composite structure. Sung H. Ko (Naval Undersea Warfare Ctr. Div., Newport, RI 02841)

A theoretical model was developed to evaluate flow noise levels received by a velocity sensor that is embedded within an outer decoupler mounted on a low acoustic impedance surface. The model, a three-layer structure, consists of a layer of nonvoided elastomer (outer decoupler), which covers a layer of microvoided elastomer (inner decoupler) backed by an elastic plate. The rubberlike material of the outer decoupler is designed to reduce the flow noise generated by turbulent boundary layer pressure fluctuations, and the low acoustic impedance (soft) material of the inner decoupler is configured to reduce the flexural wave noise generated by the vibration of the elastic plate. The upper surface of the outer decoupler is in contact with turbulent flow (water) and the lower surface of the backing plate is in contact with semi-infinite space (air). This work uses the model of the three-layer composite structure for the analyses of both the flow noise and signal levels received by the embedded velocity sensor. The final noise levels are the equivalent plane-wave levels, which were obtained by subtracting the signal levels from the flow noise levels. [Work supported by the Office of Naval Research, Code 321SS.]

3:45

4pEA5. A refined four-load method for the evaluation of in-duct acoustic source characteristics. Seung-Ho Jang and Jeong-Guon Ih (Dept. of Mech. Eng., Korea Adv. Inst. of Sci. and Technol., Science Town, Taejon 305-701, Korea)

By using the four-load method, one-port source characteristics can be determined in the duct system. In this study, a modification is made to avoid the instability problems occurring in the conventional four-load method. In the present method, the acoustic power spectra are measured at the inside or outside of the duct varying the acoustic loads, and an error function based on the linear time-invariant source model is obtained. It is numerically shown that the method is less sensitive to input errors compared to the conventional four-load method. The two methods are utilized to obtain the source parameters of a loudspeaker and a centrifugal fan using various types of loads. The estimated impedance values are compared with those obtained from the direct measurement method. As a direct measurement method, the multiple sensor method is adopted using three microphones. It is observed that the conventional four-load method results in large errors, whereas the present method yields far better agreement with the actual source characteristics.

THURSDAY AFTERNOON, 4 DECEMBER 1997 COUNCIL ROOM, 1:30 TO 5:30 P.M.

Session 4pMU

Musical Acoustics: Computer Jazz Improvisation

James W. Beauchamp, Chair

University of Illinois, Music and Electrical and Computer Engineering, 1114 West Nevada, Urbana, Illinois 61801

Chair's Introduction—1:30

Invited Papers

1:35

4pMU1. GenJam: An interactive genetic algorithm jazz improviser. John A. Biles (Information Technol. Dept., Rochester Inst. of Technol., 102 Lomb Memorial Dr., Rochester, NY 14623-5608, jab@it.rit.edu)

GenJam is an interactive genetic algorithm that learns jazz improvisation. It uses two hierarchically related populations to represent melodic ideas at the measure and phrase levels. These populations are evolved using tournament selection, single-point crossover, musically meaningful mutation, and replacement with a 50% generation gap. Fitness for the individual measures and phrases is derived from real-time feedback, which is provided by a human mentor while GenJam improvises to the accompaniment of a synthesized rhythm section. GenJam has been used for actual gigs under the billing Al Biles Virtual Quintet, which features the author on trumpet and GenJam on a variety of synthesized instruments, playing a repertoire of over 90 tunes in a variety of jazz, Latin, and new-age styles. Recent enhancements to GenJam include a pitch-to-MIDI capability that allows GenJam to listen to a human soloist, map his four-bar phrases to the GenJam genetic representation for phrases, apply selected mutation operators to these phrases, and play them back in real time as it trades fours or eights with a human soloist. In this way GenJam is truly interactive in performance, as well as during training. The lecture will feature a live demonstration of GenJam.

2:05

4pMU2. Aesthetic considerations for the composition and performance of computer-based jazz. Neil Leonard III (Music Synthesis Dept., Berklee College of Music, 1140 Boylson St., Boston, MA 02215, nleonard@it.berklee.edu)

Computers, critical doubts not withstanding, provide an important way to extend jazz improvisation and compositional resources. Jazz musicians have long explored the use of new technologies and have given a number of recent instruments their first virtuosos and repertoires. Just as jazz had given the saxophone its voice, through musicians like Coleman Hawkins, it also established the validity of the drum set in the work of Baby Dodds among others, and the electric guitar in the breakthroughs of Charlie Christian. I have sought to expand my work as a jazz improviser and composer through the use of interactive systems. For the past 12 years I have presented concerts that included original compositions for saxophone and computer controlled electronics. In the process I have explored several approaches to designing and performing with these systems, and addressed issues of how jazz is evolving. The lecture will include a demonstration of one such composition that uses an interactive system.

2:35

4pMU3. Statistical information for the development of computer models of jazz. Topi V. Jarvinen (Dept. of Musicology, Univ. of Jyvaskyla, P.O. Box 35, FIN-40351 Jyvaskyla, Finland, tjarvine@cc.jyu.fi)

Computer models of jazz improvisation have often been based on modelers' personal musical experiences or on anecdotes about jazz musicians' preferences. While this type of information is essential for any study of jazz, it would seem that it should be supplemented with some more concrete evidence on the underlying principles of jazz. This paper presents the results of a few recent statistical studies on bebop-styled jazz improvisation [e.g., T. Jarvinen, ''Tonal Dynamics and Metrical Structures in Jazz Improvisation,'' Dept. of Musicology, Univ. of Jyvaskyla, Finland]. In particular, it is discussed how bebop musicians employ meter, global tonality, and development of tonality in order to create coherent improvisations. It is argued that these types of results may help researchers design models with greater relevance and more explicit premises. An example of an application of these results is presented in another paper in this session by Petri Toiviainen.

3:05

4pMU4. Understanding and following a jazz improvisation in real time: A statistical approach. Petri H. Toiviainen (Dept. of Musicology, Univ. of Jyvaskyla, P.O. Box 35, FIN-40351, Jyvaskyla, Finland, ptoiviai@jyu.fi)

A system for following an improvised jazz solo and inferring the current location in the chord progression is presented. The system will ultimately work in real time in a MIDI environment: It listens to an improvised jazz melody played on a MIDI instrument and joins in the performance with a synthesized rhythm section. It is comprised of two stages: a beat tracker and a performance analyzer. The beat tracker is based on a continuously adapting nonlinear oscillator, which synchronizes to the beats of the improvised solo by continuously adapting its phase and period. It can follow local variations of tempo and produce expectations of next beats. The performance analyzer attempts to infer the current location in the harmonic progression by comparing the notes played in the improvised melody with statistical distributions of notes obtained from improvisations of well-known jazz musicians [T. Jarvinen, *Tonal Dynamics and Metrical Structures in Jazz Improvisation* (Dept. of Musicology, Univ. of Jyvaskyla, Finland) and trying to find the best match from several possible alternatives. [Work supported by the Academy of Finland.]

3:35

4pMU5. Representing groove: Rhythmic structure in interactive music performance. Vijay Iyer (Ctr. for New Music & Audio Technol., Univ. of California at Berkeley, 1750 Arch St., Berkeley, CA 94709, vijay@cnmat.berkeley.edu), Jeff Bilmes, Matt Wright, and David Wessel (Univ. of California at Berkeley, Berkeley, CA 94704)

The concept of ''groove,'' a musical element particularly operative in African-derived musics such as jazz, may be described as a collectively determined and relatively isochronous pulse that can be inferred from the interaction of a number of interlocking rhythmic groups. A powerful representation for musical rhythm as it appears in groove-oriented contexts has been developed. Implemented in the MAX music-programming environment, the representation encompasses pitch, accent, rhythmic deviations, tempo variation, note durations, and probabilistic processes. The design facilitates ''bottom-up'' combination techniques such as the construction of large musical objects by assembling small musical ''cells'' in series or parallel, and thus it allows improvisatory manipulation of simple rhythmic structures. The richness of control over many meaningful musical quantities distinguishes the representation from those in more common usage, such as music notation programs, sequencers, and drum machines. The implementation supports a variety of creative applications in improvised performance. Rather than attempting to make the system ''sound human'' by forcing it to simulate specific human activities, two alternative approaches are taken: the development of novel, sensitive control structures, and/or the emphasis on the system's active role as a situated participant in an interactive musical environment.

Contributed Paper

4:05

4pMU6. A novice's interface to programming digital synthesizers based on genetic algorithms. Frode E. Sandnes (Dept. Comput. Sci., Univ. of Reading, Reading RG6 6AY, England) and Craig Robertson (Univ. of Edinburgh, Edinburgh, Scotland)

The last decade has seen vast advances in commercial synthesizer developments, both in cost and features. Programming of synthesizers requires patience and detailed knowledge about sound synthesis. Also, the low-cost interfaces to these units are time consuming to access compared to the sliders of analogue synthesizers. Many musicians resort to costly off-the-shelf preprogrammed sound libraries, often without the actual characteristics and textures required. Therefore, a novel approach to programming of digital synthesizers is proposed. The task of programming is defined as an optimization problem, where the optimization function is the musician's rating of the sound, and the function arguments are the actual parameters of the sound synthesis algorithm, such as ADSR, LFO's, etc. These parameters are coded as chromosomes, and a genetic algorithm is used to search for the desired sounds. The interface works by presenting an initially random palette of sounds to the musician, for him/her to evaluate. Based on these evaluations, the genetic algorithm applies standard genetic operators to the chromosomes and over time breeds the desired sounds. Genetic algorithms ensure rapid convergence. This makes the technical details of a synthesizer transparent to the musician while it still allows him/her to explore all of its advanced features.

4:20–4:30 Break

4:30–5:30
Concert Demonstration

Session 4pNS

Noise and Engineering Acoustics: Computational Aeroacoustics

Luc Mongeau, Chair
School of Mechanical Engineering, Purdue University, 1077 Ray W. Herrick Laboratories, West Lafayette, Indiana 47907-1077

Chair's Introduction—1:00

Invited Papers

1:05

4pNS1. Recent simulations in computational aeroacoustics. Philip J. Morris (Dept. of Aerosp. Eng., Penn State Univ., University Park, PA 16802)

Computational aeroacoustics involves the direct simulation of noise production and radiation from unsteady flows. Its application to problems of practical interest has depended on both algorithm development and advances in computer technology. This paper presents some recent computational aeroacoustics calculations in several different areas. In each case parallel computations are performed to improve efficiency. A new methodology for the numerical simulation of high-speed jet noise is described. It is based on a separation of the instantaneous flow variables into time-averaged and disturbance components. The disturbances are obtained from a three-dimensional, time-dependent, compressible flow simulation. Results are presented for the effects of Mach number and temperature on jet noise. In the simulation of broadband noise from rotors and propellers, a key component is the noise generated by the interaction of a vortical gust with an airfoil. Simulations are described for such an interaction including studies of the effect of airfoil shape and loading. Finally, some examples of the simulation of acoustic scattering in both two and three dimensions from complex geometry bodies in nonuniform flow are described. A discussion of the choice of algorithms, including discretization, boundary treatments, artificial dissipation, and steady-state computations, is provided. [Work supported by NASA.]

1:35

4pNS2. A superior Kirchhoff method for aeroacoustic noise prediction: The Ffowcs Williams–Hawkings equation. Kenneth S. Brentner (NASA Langley Res. Ctr., Hampton, VA 23681, k.s.brentner@larc.nasa.gov)

The Lighthill acoustic analogy, as embodied in the Ffowcs Williams–Hawkings (FW–H) equation, is compared with the Kirchhoff formulation for moving surfaces. A comparison of the two governing equations reveals that the main Kirchhoff advantage (namely, nonlinear flow effects are included in the surface integration) is also available to the FW–H method if the integration surface used in the FW–H equation is not assumed impenetrable. The FW–H equation is analytically superior for aeroacoustics because it is based upon the conservation laws of fluid mechanics rather than the wave equation. Hence, the FW–H equation is valid even if the integration surface is in the nonlinear region. This is demonstrated numerically for helicopter rotor applications in the paper. The Kirchhoff approach can lead to substantial errors if the integration surface is not positioned in the linear region (i.e., if the input data are not a solution to the wave equation). These errors may be hard to identify in some cases.

Contributed Paper

2:05

4pNS3. Large-eddy simulation of compressible free jet turbulence applied to computation of exhaust mixing noise. David B. Schein (Northrop Grumman Corp., 9H11/GK, 8900 E. Washington Blvd., Pico Rivera, CA 90660 and Dept. of Mech. and Aerosp. Eng., UCLA) and William C. Meecham (UCLA, Los Angeles, CA, meecham@seas.ucla.edu)

A computational study of free, heated jet flow and resultant far-field sound was performed using large-eddy simulation (LES) and Lighthill's acoustic analogy. A subgrid scale model for small-scale compressible turbulence was developed using a combination of the popular Smagorinsky model and a deductive model. The primary objective is to address large Reynolds number (Re), high subsonic (compressible) flow with realistic geometries more representative of aircraft engine exhausts than typically considered using direct numerical simulation (DNS). Flow field fluctuations are stored over a period of time and used to calculate rms turbulence within the computational domain. The far-field sound and directivity is computed using the time-derivative form of Lighthill's source–integral result formulated in terms of quadrupole sources from the simulated flow field, which is integrated in time and contains the fluctuations set up by the time-varying stress tensor. A simulation for a WR19-4 turbofan engine exhaust ($Re \approx 2 \times 10^6$ based on exit velocity and diameter) is presented, and propagated jet noise results are compared with experimental acoustics data. Methods to account for effects of source convection and thermal refraction to obtain realistic frequency spectra and directivity are considered.

Session 4pPA

Physical Acoustics: Bubbles, Drops and Particles

Robin O. Cleveland, Chair
Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Contributed Papers

1:30

4pPA1. Mutual interaction of a two-phase turbulent/bubbly submerged water jet with a single-phase jet for the measurement of the enhanced hydrodynamic near-field pressure spectrum. Murray S. Korman (Dept. of Phys., U.S. Naval Acad., Annapolis, MD 21402)

The interaction of mutually perpendicular submerged turbulent water jets is studied from the measurement of the enhanced hydrodynamic near-field pressure spectrum. One jet is a conventional single-phase, free turbulent shear flow circular jet, while the other jet is a two-phase bubbly jet of similar construction. The overlap region is located at four-nozzle diameters (4D) away from each orifice. A small hydrophone located outside the overlap region (4D away from the interaction region and 45 deg from the jet axis) measures the near-field pressure spectrum. Very small gas bubbles are required for this two-phase turbulent jet. They are generated by pressurized carbonated water flowing through a thin circular nozzle plate having an array of small holes. These CO_2 gas bubbles are mixed within the entrance to the tapered jet nozzle. Alternatively, bubbles are generated by pressurized N_2 gas passing through a fritted disk housed in a Buchner funnel located near the nozzle entrance. Spectra are measured and compared for different void fractions and different Reynolds numbers that are based on the jet diameter.

1:45

4pPA2. Acoustic microcavitation at surfaces: New results in deinking of paper. Sameer I. Madanshetty (Mech. and Nuclear Eng., Kansas State Univ., 339 Durland Hall, Manhattan, KS 66506-5106)

Deinking remains an important step in environmentally conscious manufacturing of paper. A novel method based on acoustic microcavitation is used for deinking xerographic ink from paper. Microcavitation evolves microimplosions, which are effective in causing deinking preferentially at ink sites. Acoustic microcavitation is brought about by low megahertz acoustic fields giving rise to micron-sized bubbles that live a few microseconds. In exposing a surface to continuous waves for a defined duration one could obtain cavitation effects in an average, overall sense; the details of nucleation, evolution of inertial events, and the precise interplay of field parameters in effecting cavitation, however, get glossed over. Studying pulsed cavitation using tone bursts at low duty cycles, instead of CW insonification, reveals interesting details of the initiation and evolution of acoustic microcavitation. Experiments indicate that the 2-D context of how a xerographic print is deinked might be useful in inferring some attributes of 3-D pulp deinking. Assuringly, the results indicate that paper fibers are entirely undamaged, and the ink separation leaves them immaculately white, and that acoustic methods for deinking are a viable, chemical free, environmentally responsible means of recycling paper.

2:00

4pPA3. Acoustic radiation force on micrometer-size particles: Size dependency and the efficiency of the superposition method. Kenji Yasuda (Adv. Res. Lab., Hitachi Ltd., 2520 Akanuma, Hatoyama, Saitama 350-03, Japan)

The acoustic radiation force on polystyrene spheres was measured through observation of the sphere movement in a 500-kHz ultrasonic standing wave. As predicted by Yosioka, the linear dependency of the force on the cube of the sphere radius began to fail when the sphere radius was below 5 μm. This failure can be accounted for by the presence of a shell layer that increased the effective radius of the sphere and it suggests that acoustic radiation might be used in handling microspheres smaller than previously thought [K. Yasuda and T. Kamakura, Appl. Phys. Lett. (in press)]. A method for concentrating particles using an acoustic radiation force by the superposition of the higher harmonics on fundamental ultrasound has also been investigated. The efficiency obtained when blood cells are concentrated by using the superposition method should theoretically be 20% better than that obtained when using the sine wave. Although the improvement obtained experimentally was only 5% because the waveform was incomplete, the potential usefulness of the superposition method has been demonstrated. [K. Yasuda, Jpn. J. Appl. Phys. **36**, 3130–3135 (1997)].

2:15

4pPA4. Novel configurations for acoustophoresis. Todd L. Brooks and Robert E. Apfel (Yale Univ., New Haven, CT 06520-8286)

Separation of particles by their acoustomechanical properties can be accomplished by acoustophoresis, which depends on both primary and secondary acoustic radiation forces. In an acoustic standing wave, particles are moved either toward the pressure nodes or the antinodes depending on the relative contrasts in the particle density and compressibility with the surrounding host liquid. Because particles reradiate the sound waves, there are secondary interparticle forces as well. These secondary forces can either hinder or enhance the separation process. If two species which are to be separated attract each other because of interparticle forces, the separation will be compromised. But if several like particles aggregate, then primary forces are much stronger and the separation occurs more rapidly. When flow is superimposed on the system, as is common in practical applications, then drag forces must also be accounted for. In the present work, novel field configurations have been explored, and the trajectories of particles have been computed for a number of frequencies, flow speeds, field strengths, and property contrasts. Based on these simulations, experiments have been designed to optimize the efficiency of acoustophoresis, especially for biotechnology applications. [Work supported by NASA through Grant NAG8-1351.]

2:30

4pPA5. Numerical simulation of superoscillations of a BSA-bearing drop in microgravity. Xiaohui Chen, Robert E. Apfel (Dept. of Mech. Eng., Yale Univ., 9 Hillhouse Ave., New Haven, CT 06520-8286), and Tao Shi (Thomas Jefferson Univ., Philadelphia, PA 19107)

Large-amplitude nonlinear oscillations of a water drop, radius 2.53 cm and aspect ratio 2.10, with surfactant BSA (bovine serum albumin) 1.0×10^{-5} g/ml(1CMC), is numerically simulated based on the boundary integral method. BSA is a globular, large molecular-weight protein. Two surface viscosities, surface dilatational viscosity and surface shear viscosity, as well as the Marangoni effect, are considered. The high-surface viscosities make the rotational velocity the same order as the irrotational velocity in the boundary layer. A boundary-layer method is used to calculate the rotational velocity. When the surface dilatational viscosity is 0.5 sp and the surface shear viscosity is 0.15 sp, the numerical simulation results are in good agreement with the experiment results observed in the space shuttle during the second United States Microgravity Laboratory, USML-2, in October, 1995. The evolution of the drop oscillation for both experiment and simulation is given. The change of the BSA distribution along the surface, and the relative importance of the rotational and irrotational velocities, as the drop oscillates is also shown. [Work supported by NASA through JPL, Contract No. 958722.]

2:45–3:00 Break

3:00

4pPA6. Drop and bubble dynamics investigations on Earth and in low gravity using ultrasonics. Eugene H. Trinh (JPL/Caltech, MS 183-401, 4800 Oak Grove Dr., Pasadena, CA 91109)

Two ultrasonic devices for the positioning and the remote manipulation of free drops and bubbles have been used during the recent STS-94 Spacelab flight. One apparatus is designed to levitate or to position single or a small number of individual droplets in air, to induce drop shape oscillations, and to control the residual drop rotation. Implementation in low gravity has allowed the measurement of droplet dynamics and rotational stability in the absence of the overwhelming constraint imposed by the high-intensity ultrasonic field required for levitation on Earth. The second apparatus permits the trapping or positioning of gas bubbles in a water-filled resonant cell with a square cross section. Confirming earlier results from a previous flight experiment, it was found that gravity plays a determining part in the stable centering of bubbles larger than resonant size. Both devices operate at about 22.5 kHz and allow the monitoring of the drop or bubble motion through the detection of scattered light from a collimated diode laser beam illuminating the fluid particles. These investigations are low cost, they are built from commercially available components, and they are manually operated by crew members in the Middeck/Spacelab Glovebox facility. [Work funded by NASA.]

3:15

4pPA7. Single-bubble sonoluminescence in microgravity. Thomas J. Matula, Jarred E. Swalwell, Vassiloius Bezzerides, Paul Hilmo, Mike Chittick, Lawrence A. Crum (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105), David W. Kuhns (Univ. of Washington, Seattle, WA 98105), and Ronald A. Roy (Boston Univ., Boston, MA 02215)

A recent experiment involving single-bubble sonoluminescence aboard NASA's parabolic research aircraft will be described. Measurements of the intensity of the light emission were performed during periods of microgravity (near 0 g) and hypergravity (near 2 g). Gravitational effects on the luminescence and extinction pressure thresholds were examined. In addition, the light emission was monitored under constant drive conditions as the gravitational acceleration varied during the parabolas. Measurements show some variability in the sonoluminescence intensity for both thresholds. However, during otherwise constant conditions in which the drive pressure amplitude did not change, the intensity from stable single-bubble sonoluminescence was observed to increase during periods of microgravity. There was an initial increase in intensity that occurred simultaneous with the decrease in the gravitational acceleration, followed by a slow increase in intensity that appeared to level off near the end of the microgravity period. The longer time scales over which the intensity changed during the microgravity period may be indicative of gas diffusion occurring as the bubble attains a new equilibrium condition. [Research supported by NSF and NASA.]

3:30

4pPA8. Measurements of the transient response of single-bubble sonoluminescence subject to an abrupt change in the drive pressure amplitude. Thomas J. Matula and Lawrence A. Crum (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

A novel method utilizing abrupt drive pressure amplitude changes is described for exploring the dynamics of single-bubble sonoluminescence. Air bubbles that are initially below the luminescence threshold are subject to an abrupt increase in the drive pressure amplitude. The resulting turn-on time for light emission is qualitatively similar to nitrogen bubbles, except that over time scales of seconds, the light intensity of an air bubble increases [B. P. Barber *et al.*, ''Sensitivity of sonoluminescence to experimental parameters,'' Phys. Rev. Lett. **72**, 1380 (1994)], while the intensity from a nitrogen bubble remains low. However, when we begin with a stable sonoluminescing air bubble *above* the luminescence threshold, and then generate an abrupt decrease in the pressure amplitude below the threshold (no light regime), followed again by an abrupt increase back to its original (sonoluminescing) state, the resulting turn-on time for light emission appears qualitatively similar to noble gas bubbles: sudden, and relatively intense, unlike what occurs for air bubbles that were initially below the luminescence threshold. These measurements imply that sonoluminescence from an air bubble depends on the time the bubble spends in the sonoluminescing state. [Research supported by NSF.]

3:45

4pPA9. Could optical radiation pressure be used to move a sonoluminescing bubble to the pressure antinode of an acoustic standing wave? David B. Thiessen and Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

It is generally recognized that in SBSL, the bubble is displaced slightly above the antinode of the acoustic standing wave which drives the bubble oscillations. This is because the acoustic radiation force must balance the average buoyancy of the bubble. For scientific purposes and to simulate SBSL in microgravity, it would be desirable to reposition the bubble at the antinode. One strategy to achieve that repositioning where the buoyancy force would be balanced entirely by the average optical radiation pressure of a downward propagating laser beam has been numerically investigated. Optical levitation of small bubbles has been previously demonstrated [B. T. Unger and P. L. Marston, J. Acoust. Soc. Am. **83**, 970–975 (1988)]; however, it is not trivially applied to SBSL; with cw illumination of a quiescent bubble, the laser power required increases as the cube of the bubble radius. To simulate the large-radius phase of SBSL bubble oscillations, a modified Rayleigh–Plesset equation was numerically integrated. The required power for cw and pulsed laser beams was estimated. The average power is significantly reduced if the beam is pulsed to coincide with the large-radius phase of the oscillations so as to maximize the relevant optical cross section.

4:00

4pPA10. Numerical research of the influence of external conditions on the sonoluminescence process. Vladimir N. Nogin, Nikolay G. Karlykhanov, Gennady V. Kovalenko, Vadim A. Simonenko (RFNC-VNIITF, Snezhinsk, Russia), and William C. Moss (Lawrence Livermore Natl. Lab., Livermore, CA)

Using a 1-D approximation, processes are numerically simulated that take place during oscillations of a single gas bubble located in the center of a spherical flask with liquid exposed to a periodical acoustic wave. Real equations of state for water, gases (air, argon), surface tension, and heat conduction (molecular and radiative) are taken into account in the simu-

lation. The role of shock-wave processes at bubble compression is investigated. Because of the peculiarities in the behavior of the gas, the equation of state is shown to be the most important at densities greater than 1 g/cm^3, with the parameters of bubble compression being weakly affected by the water equation of state. Time dependencies of the bubble radius during one acoustic cycle are obtained for different ambient conditions: acoustic-wave amplitude and frequency, liquid temperature, and magnitude of surface tension. Comparison with experimental data is performed. [Work supported by CRDF.]

4:15

4pPA11. Scaling analysis for single-bubble cavitation. Vadim A. Simonenko, Nikolay G. Karlykhanov, Gennady V. Kovalenko, Vladimir N. Nogin (RFNC-VNIITF, Snezhinsk, Chelyabinsk Region, Russia), and William C. Moss (Lawrence Livermore Natl. Lab., Livermore, CA)

There is a new interest to study caviation phenomena for gas-filled bubble in liquid basing on recent experimental and theoretical results on single-bubble sonoluminescence. Scaling analysis is applied to these phenomena. It allows a role of main control parameters to be clarified which have an influence on energy transfer to central gas-filled cavity and on energy dissipation in gas. The main parameters are linear dimensions, initial equilibrium pressure, initial temperature, rate and amplitude of pressure change. Of principle importance are convergent symmetry parameters and time-shape for external pressure. The most important are phenomena which provide maximum energy transfer to a central bubble, and to internal gas layers in the bubble. Factors decreasing efficiency of energy transfer are shown. The main factors are (1) compressibility of liquid caused by additional energy dissipation in layers adjacent to bubble; (2) early origin of shock wave near gas–liquid boundary. Theoretical estimations are confirmed by simulations calibrated on single bubble sonoluminescence. Criteria are shown to optimize gas compression in bubble. There are proposed prospective experiments to study large-scale effects in systems with external size of few meters and equilibrium pressure of several kilobars. [Work supported by the CRDF.]

THURSDAY AFTERNOON, 4 DECEMBER 1997 CHAMBER ROOM, 2:00 TO 4:45 P.M.

Session 4pPP

Psychological and Physiological Acoustics: Psychophysical Beachcombing

Kourosh Saberi, Chair
Division of Biology, Caltech, Pasadena, California 91125

Contributed Papers

2:00

4pPP1. First and second pitch shift effects as alternative manifestations of a single phenomenon. Pantelis N. Vassilakis (MA Ethnomusicol., Dept. of Ethnomusicol., Program in Systematic Musicol., 2539 Schoenberg Hall, UCLA, Los Angeles, CA 90095, pantelis@ucla.edu)

The second pitch shift effect describes a drop in the perceived pitch of a complex stimulus when the frequency spacing among its components is increased. Based on existing experimental data describing the first pitch shift effect, the present paper demonstrates that the mathematical model introduced by de Boer [Doctoral dissertation, University of Amsterdam (1956)] and modified later by Smoorenburg [J. Acoust. Soc. Am. **48**, 1055–1060 (1970)], predicts without any further modification the second pitch shift effect as well. The Smoorenburg model also predicts that the perceived pitch of complex stimuli will not always drop when increasing the frequency spacing among components, but may rise depending on the structure of the stimuli. A perceptual experiment was conducted using nine complex stimuli. For eight of the stimuli, the model predicted a rise in pitch with increasing frequency spacing while for the ninth stimulus the opposite was predicted. The results of the experiment support the conclusion that the relationship between the direction of pitch motion (rise/drop) and the direction of changes in frequency spacing among components of complex stimuli (increasing/decreasing) can be predicted by the same model that explains the first pitch shift effect, making the second pitch shift effect an unnecessary concept.

2:15

4pPP2. Psychometric functions for two-transient dichotic stimuli. Agavni Petrosyan (Dept. of Psych., California State Univ., Los Angeles, CA 90032) and Kourosh Saberi (Caltech, Pasadena, CA 91125)

To better understand the parameters that limit interaural-delay sensitivity in a precedence-effect task, psychometric functions were measured for the detection of an interaural time difference (ITD) in the second transient of a two-transient binaural stimulus. The stimuli were broadband 1-ms Gaussian pulses centered on the cosine phase of a 6-kHz carrier. Twenty-point psychometric functions were obtained using a 2IFC method of constant stimuli, with each point based on a minimum of 500 trials. Four functions were obtained for each observer at interclick intervals (ICI) of 3, 6, and 12 ms, in addition to a single-transient control condition. Proportion of correct responses were transformed to d′ units and linearly fitted with $\ln(d') = \ln\beta + \upsilon \ln(\Delta\mathrm{ITD})$, where $\ln\beta$ and υ are the parameters of the fit. The slope parameter υ was independent of ICI and remained near unity, discounting nonlinear transformations of the stimulus scale [J. P. Egan, J. Acoust. Soc. Am. **38**, 1043–1049 (1965)]. However, $\ln\beta$ decreased with decreasing ICI, suggesting linear changes in either the variance or expected value of the perceived lateral position. Further experiments examined the contribution of each of these latter factors to changes in $\ln\beta$. Implications for causes that underlie the precedence effect are discussed.

2:30

4pPP3. The effects of nonconcurrent tonal and noise distractors upon auditory spatial resolution: The temporal, spatial, and spectral parameters underlying the ''disruption effect.'' Eric A. Erpenbeck, David R. Perrott, and Raymond Lim (Dept. of Psych., California State Univ., 5154 State University Dr., Los Angeles, CA 90032)

The minimum audible angle (MAA) thresholds obtained by Mills [A. W. Mills, J. Acoust. Soc. Am. **30**, 237–246 (1958)] involved two 1000-ms sinusoidal tone pulses presented sequentially with an interstimulus interval (ISI) of 1000 ms. In contrast to the generally good localization performance obtained by Mills, localization of two simultaneous events (concurrent localization or CMAA) results in relatively poor performance [D. R. Perrott, J. Acoust. Soc. Am. **76**, 1704–1712 (1984)]. The current set of experiments attempts to bridge the gap between these two extreme paradigms. While the target tones (14-ms, 1000-Hz sinusoids) were presented sequentially (an MAA task), an attempt was made to systematically examine the consequences of the addition of a nonoverlapping, auditory

event (tones or noise) during the interval between the target events. The results of this work indicate that localization performance is extremely sensitive to events that occur between the target events. The disruptive effect of a ''nontarget'' event during the ISI could even be observed when relatively long ISIs (814-ms) and very brief (14-ms) distractors events were used. Interesting interactions were encountered as a function of the spatial positioning of the nontarget (''distractor'') event. Implications of these results will be discussed.

2:45

4pPP4. Individual subjective preference of simulated sound fields by listeners for opera sound sources in relation to the subsequent reverberation time. Hiroyuki Sakai (Grad. School of Sci. and Technol., Kobe Univ., Rokkodai, Nada, Kobe, 657 Japan), Hiroshi Setoguchi (Miyama Conceru, Aira, Kagoshima, 899-66 Japan), and Yoichi Ando (Kobe Univ., Rokkodai, Nada, Kobe, 657 Japan)

The purpose of this study is to evaluate the subjective preference of a simulated sound field by listeners in changing subsequent reverberation time T_{sub} using vocal sources. A great deal of effort has been made in the study using music or speech [for example: Ando *et al.*, Acustica **50**, 134–141 (1982); Ando *et al.*, J. Acoust. Soc. Jpn. **39**, 89–95 (1983), in Japanese]. Subjective evaluation using vocal sources has never been examined, although vocal performance is often played in concert halls and opera houses. Subjective preference tests were conducted changing T_{sub}, which is one of the four objective parameters in relation to subjective preference of sound field. Individual differences of subjective preference as well as global preference are also considered.

4pPP5. Abstract withdrawn

3:00

4pPP6. Effects of changes of the spectral masking slope on sound quality and clarity of music sounds in the normal and impaired ear. Bernhard Laback (Acoust. Res. Lab., Austrian Acad. of Sci., Liebigg. 5, A-1010 Wien, Austria), Niek Versfeld (Free-Univ.-Hospital, 1007 MC Amsterdam, The Netherlands), and Werner Deutsch (Austrian Acad. of Sci., Liebigg. 5, A-1010 Wien, Austria)

Reduced frequency selectivity as occurs in the cochlear impaired ear has been shown to be a main factor for degraded speech perception [J. M. Festen and R. Plomp, J. Acoust. Soc. Am. **76**, 652–662 (1983)]. It is also assumed to have negative effects on music perception. In this paper the hypothesis is tested that the loudness relations between spectral peaks and nonmasked lower-level components, which depend on the individual frequency selectivity, are important for (a) subjective sound quality (listening comfort) and (b) clarity, as measured in terms of detectability of altered notes in music excerpts. In order to test this hypothesis a signal processing algorithm has been utilized which enhances or suppresses the lower-level spectral components of a complex signal according to a masking function. Processed signals have been presented to normal-hearing and cochlear-impaired subjects. Results revealed that subjects with reduced frequency selectivity—as measured with psychoacoustical tuning curves—tend to prefer music signals with enhanced lower-level components. In the clarity experiment, results obtained with hearing-impaired subjects show improved detectability with suppression of the lower-level components (increase of spectral contrasts). [Work supported by Austrian Academy of Sciences.]

3:15–3:30 Break

3:30

4pPP7. Evaluation of simulations of the effects of hearing loss produced by combinations of dynamic expansion and spectral smearing. Isaac J. Graf, David S. Lum, and Louis D. Braida (Res. Lab. of Electron., MIT, Cambridge, MA 02139)

By processing sounds with combinations of multiband expansion (MBE) and spectral smearing (SS), listeners with normal hearing can experience the elevated thresholds, abnormal growth of loudness, and reduced frequency selectivity similar to those of listeners with sensorineural hearing impairments. The effects of simulations of flat 40-dB losses produced by such combinations on masking patterns and loudness summation, and of simulations of a more severe loss on speech perception in noise, were studied. Both MBE and SS reduce frequency selectivity as measured by masking patterns for tones produced by narrow-band noise, and both broaden forward-masked tuning curves for tones. However, for equally loud probe tones, MBE elevates the tips of psychoacoustic tuning curves much more than SS. As measured by loudness balances between narrow-band and wideband noises, MBE reduces loudness summation much more than SS. The effects of MBE and SS simulations of a severe hearing loss on consonant identification in speech-shaped noise depend on the SNR, with MBE more detrimental at high SNR and SS more detrimental at low SNR. MBE tends to degrade reception of cues for voicing; SS, cues for place of articulation; both degrade cues for manner of articulation. [Work supported by NIH.]

3:45

4pPP8. A psychophysically based model of consonant perception by multichannel cochlear implant users. Ted A. Meyer, Mario A. Svirsky, Michael B. Castor, Sassan Falsafi (Dept. of Otolaryngol., Indiana Univ. School of Medicine, 702 Barnhill Dr., Indianapolis, IN 46202), and Peter M. Simmons (Duke Univ., Durham, NC)

Despite advances in implant technology, cochlear implant (CI) users demonstrate a wide range in the ability to perceive speech in the absence of visual or contextual cues. Some progress has been made recently in the ability to explain perceptual performance with a CI. A quantitative, psychophysically based model of vowel perception by CI users of the SPEAK processing strategy generated confusion matrices that were remarkably similar to actual data [M. A. Svirsky and T. A. Meyer, Assoc. Res. Otolaryngol. Abs. **20**, 59 (1997)]. The three dimensions of the parameter space are the centers of gravity of stimulation in the cochlea in response to the first three vowel formants. The free parameter of the model is the subject's ability to scale pitch percepts associated with different electrodes. Although this single-parameter model was not successful in predicting consonant confusions, when the model was expanded to incorporate psychophysical estimates of gap detection and high- versus low-frequency intensity difference discrimination, the model successfully predicted performance on a medial consonant test [Meyer *et al.*, Conf. Implant. Aud. Prosthes. Abs. (1997)]. Relations between predicted and obtained performance on selected psychophysical tests will be discussed. [Work supported by NIH.]

4:00

4pPP9. Discrimination of single-formant stimuli by chinchillas (*Chinchilla villidera*). Lori L. Holt, Keith R. Kluender (Dept. of Psych., Univ. of Wisconsin, 1202 W. Johnson St., Madison, WI 53706), and Andrew J. Lotto (Loyola Univ., Chicago, IL 60625)

The present study assessed the behavioral capacity of six chinchillas (*Chinchilla villidera*) to discriminate complex acoustic differences, like those observed for the first formant ($F1$) of human speech, using operant methods. In a minimum-uncertainty task, chinchillas discriminated single-formant stimuli with fundamental frequencies of 100 or 200 Hz, 60 Hz ($Q6$) bandwidths, and center frequencies ranging from 2.5–7.0 Bark (approximately 253–709 Hz) in 0.25-Bark steps. Each trial began with the repeated presentation of a ''target'' stimulus. The chinchillas' task was to discriminate whether this repeating target changed in center frequency or remained constant throughout the trial. The chinchillas completed 20 experimental sessions for each of six conditions in which fundamental frequency and center frequency of the target stimulus were varied. Stimulus characteristics were designed to reveal perceptual consequences of interactions between individual harmonics and overall spectral shape of the single formant. Behavioral sensitivities were estimated by d′ and percent correct. Data will be discussed in the context of human psychophysical performance and observations of single-unit responses in the chinchilla cochlear nucleus from studies utilizing the same stimulus corpus. [Work supported by NSF.]

4:15

4pPP10. Auditory processing is related to reading ability. Athanassios Protopapas, Merav Ahissar, and Michael M. Merzenich (Sci. Learning Corp., 417 Montgomery St., Ste. 500, San Francisco, CA 94104)

Aspects of auditory processing related to temporal and spectral resolution were investigated in 50 adult subjects (aged 18-58) with varying reading abilities. The tests administered included tone detection, tone frequency discrimination, tone sequencing, interval discrimination, and gap detection. Tones were either long (250 ms) or short (20 ms) and ranged in frequency between 600 and 1400 Hz. Detection and frequency discrimination tasks were given in the clean and in a masking context, in which tones were followed by 300-ms bandpass noise. Thresholds were determined using an adaptive procedure. Backward masking interferences for short tones, as well as interval and frequency discrimination limens were substantially elevated for most poor readers, and were correlated with one another. Three-tone sequencing and short-tone frequency discrimination in a masking context were especially strongly correlated with single-word reading ability. Gap detection thresholds did not correlate with other tasks or with reading ability. In sum, spectral-temporal auditory processing resolution appears to be related to reading ability, possibly via an acoustically based deficit in phonetic development. These findings raise the question of whether nonspeech acoustic training can be used to improve adults' reading ability. [Work supported by Scientific Learning Corporation.]

4:30

4pPP11. On the apparent source width for music sources related to the IACC and the width of the interaural crosscorrelation function (W_{IACC}). Shin-ichi Sato and Yoichi Ando (Grad. School of Sci. and Technol., Kobe Univ., Rokkodai, Nada, Kobe, 657 Japan)

It has been shown that the ASW for 1/3 octave band noise can be described as functions of the IACC and the W_{IACC}, which is defined by the time interval between values of 10% below the IACC [S. Sato and Y. Ando, J. Acoust. Soc. Am. **100**, 2592 (1996)]. In this study, it is examined whether the ASW for music sources also can be calculated with the IACC and the W_{IACC}. In order to control of both the IACC and the W_{IACC}, the delay time of two reflections relative to the direct sound was changed in the range of 0–50 ms. Results of the scale value of ASW obtained by paired comparison are compared with calculated ASW.

THURSDAY AFTERNOON, 4 DECEMBER 1997 PRESIDIO ROOM, 1:00 TO 4:45 P.M.

Session 4pSC

Speech Communication: Phonetic Perception and Word Recognition

Lynne E. Bernstein, Cochair
Spoken Language Processing Laboratory, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

Edward T. Auer, Jr., Cochair
Spoken Language Processing Laboratory, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

Invited Papers

1:00

4pSC1. Chairs' introduction to special session on phonetic perception and word recognition. Edward T. Auer, Jr. and Lynne E. Bernstein (Spoken Lang. Processes Lab., House Ear Inst., 2100 West Third St., Los Angeles, CA 90057)

A classic statement of the speech perception problem is how to account for the perception of phonemes based on the acoustic stimulus. A classic statement of the word recognition problem is how to account for the access and selection of lexical representations, with little concern for phoneme/phonetic perception. Over the past two decades, research in spoken-word recognition has begun to explicitly model the stages of word recognition in terms of phonetic/phonemic perception and later stages involving lexical knowledge. These different models incorporate very specific proposals for the interaction of information from the bottom-up stimulus and the lexicon. They also raise issues concerning the appropriate model for phoneme perception, that is, for example, does it precede or follow the lexical processing stage? It would be convenient if theories/models of phoneme perception could ignore effects of the lexicon, and therefore, autonomously account for the ''front-end'' of word recognition. But can they, and be adequate theories/models

of either phoneme perception or the front-end to word recognition? Alternatively, are there phoneme identification effects that are irrelevant to word recognition? In this session, researchers representing a variety of perspectives on the true relationship between lexical processes and phoneme/phonetic perception are brought together.

1:10

4pSC2. The beginnings of word recognition in infancy. Peter W. Jusczyk (Dept. of Psych. and Dept. of Cognit. Sci., Johns Hopkins Univ., Baltimore, MD 21218, jusczyk@jhu.edu)

At around 7.5 months, English-learning infants begin to show some ability to segment words from fluent speech. Moreover, there is evidence that infants at this age begin to engage in some long-term encoding and retention of information about the sound patterns of words that occur frequently in fluent speech. Infants' memory for these sound patterns apparently does not depend on their pairing with any concrete referents. Thus, at least in some instances, word learning may occur by first storing information about sound patterns, and then subsequently attaching a meaning to these. Additional findings from our laboratory suggest that these early representations of the sound patterns of words are rather detailed with respect to phonetic properties and may include information about such indexical properties as talker identity. [Work supported by NICHD and NIMH.]

1:35

4pSC3. Spoken-word recognition. Cynthia M. Connine (Dept. of Psych., Univ. of Binghamton, SUNY, Binghamton, NY 13902-6000)

Spoken-word recognition is an efficient and generally error-free process that occurs under a variety of speaking and listening conditions. The talk will focus on the mapping process between the speech signal and access of form and meaning. The nature of the representation that supports spoken-word recognition will be discussed with a focus on the consequence of ambiguity and mismatching information. Research has been conducted in the past few years suggesting that activation of lexical representations is accomplished via feature mapping. It is argued that this architecture permits lexical activation given incomplete or erroneous input. Phonological variation and some recent work concerning representation and processing of common variants will also be discussed.

2:00

4pSC4. A comparison of perceptual word similarity metrics. Paul Iverson, Edward T. Auer, Jr., and Lynne E. Bernstein (Spoken Lang. Processes Lab., House Ear Inst., 2100 W. 3rd St., Los Angeles, CA 90057)

Contemporary theories of spoken-word recognition rely on the notion that the stimulus word is mapped against, or selected from, words in long-term memory in terms of its phonetic (form-based) attributes. A few metrics have been proposed to model the form-based similarity of words, including an abstract phonemic metric computed directly on the lexicon (i.e., Coltheart-n), and perceptual metrics derived from the results of phoneme identification experiments. The results of applying several different metrics to phoneme and word identification data (open-set and forced-choice tasks) will be discussed, and these metrics across stimulus conditions with a range of intelligibility levels and similarity structures (visual-only lipreading, audio-only conditions processed by a vocoder, and audiovisual conditions pairing vocoded audio with lipreading) will be compared. The our results suggest that graded perceptual metrics may be most useful for understanding the results of word identification experiments across a wide range of stimulus conditions. [Work supported by NIH DC00695.]

2:25–2:40 Break

2:40

4pSC5. The relation between phoneme reception and word perception. Guido F. Smoorenburg (Hear. Res. Lab., Dept. of Otolaryngol., Univ. Hospital Utrecht F.02.504, P.O. Box 85500, 3500 GA Utrecht, The Netherlands)

Word perception can be modeled as a two-stage speech recognition process involving, first, recognition of individual phonemes and, second, substitution of missing elements using contextual information. Within this conceptual framework, contextual information may be acoustic in nature, coarticulation, and it may be related to lexical context. Using nonsense CVC syllables, we have analyzed the effects of coarticulation; using meaningful CVC syllables, the additional effects of lexical context. Supplementary estimates of the effects of lexical context were obtained from an incomplete orthographic presentation of the meaningful CVC syllables. The acoustic materials were presented to normal-hearing subjects and to hearing-impaired subjects. The results show large differences in recognition probability of the individual phonemes at low speech-to-noise ratios and marked contextual effects, attributed to coarticulation, for nonsense CVC syllables presented in noise. In the Dutch language, meaningful CVC syllables appear to consist of 2.5 statistically independent elements (rather than 3), which is attributed to the lexical context effects. Near threshold, response behavior of the hearing impaired is heavily based on lexical information, which lowers this number of statistically independent elements to a value close to 1. [This research was performed in collaboration with Dr. A. J. Bosman and Dr. A. W. Bronkhorst.]

3:05

4pSC6. Lexical neighborhood and phoneme frequency effects in phonetic processing. James R. Sawusch (Psych. Dept., Park Hall, Univ. of Buffalo, SUNY, Buffalo, NY 14260, jsawusch@acsu.buffalo.edu)

The phonetic coding of speech reflects both a stimulus driven process and influences from the mental lexicon. There are at least three distinct forms of influence of the mental lexicon on phonetic processing: lexical status (is the item a word), lexical neighborhood (how many words is the item similar to), and phoneme frequency (how often does the phoneme occur in the language). Using natural speech series, the identification of an ambiguous phoneme is influenced by both the lexical neighborhood of the stimulus (syllable), and the probability of occurrence of the target phonemes in the language. Listeners consistently identify ambiguous phonemes with

the label that corresponds to the end of the series with a larger lexical neighborhood, and listeners are biased to report more probable phonemes. Neighborhood and phoneme frequency effects are also found in a lexical decision task for clear, unambiguous tokens. Together with other work on higher-level influences on phoneme perception, these results indicate that phoneme perception is the result of an interaction between acoustic–phonetic information and the lexical knowledge of the individual, regardless of whether the target item is a word or not. [Work supported by NIDCD Grant No. R01 DC00219 to SUNY at Buffalo.]

3:30

4pSC7. Transitional probability, not lexical knowledge, influences compensation. James M. McQueen (Max-Planck-Inst. for Psycholinguist., Wundtlaan 1, 6525 XD Nijmegen, The Netherlands) and Mark A. Pitt (Ohio State Univ., Columbus, OH 43210)

The perceptual system compensates for fricative–stop coarticulation. Ambiguous stops between /t/ and /k/ tend to be heard as /k/ after /s/-final words and as /t/ after /S/-final words, even when the word–final fricatives are replaced with an ambiguous phoneme [J. L. Elman and J. L. McClelland, J. Mem. Lang. **27**, 143–165 (1988)]. Biases in the transitional probabilities (TPs) of /s/ and /S/ in the words in that study mean that the effect following ambiguous fricatives could be due either to lexical involvement in prelexical compensation for coarticulation or to a sensitivity to TP at the prelexical level. In three experiments, listeners categorized nonword–final fricatives with TP biases, word–final fricatives where TPs were controlled, and word–initial stops following both fricative contexts. Categorization of ambiguous fricatives and stops was influenced by the TPs of the fricatives in the nonwords. But although there was a lexical bias in the categorization of the word–final ambiguous fricatives, there was no lexical influence on the categorization of the stops. These results suggest that the lexicon does not influence prelexical processing, and that the prelexical level is sensitive to TP. [Work supported by NIDCD, the Human Frontiers Science Project, and the Max Planck Society.]

3:55

4pSC8. Predicting the perception of words from the perception of phonemes. Terrance M. Nearey (Dept. of Linguist., Univ. of Alberta, Edmonton, AB T6G 2E7, Canada)

The present paper explores the degree to which the perception of CVC syllables can be factored into the perception of its constituent phonemes. Categorization experiments with synthetic speech typically manifest such factorability to a remarkable degree (regardless of lexical status of the syllables.) Such factorability is also compatible with intelligibility studies of English CVCs [e.g., A. Boothroyd and S. Nittrouer, J. Acoust. Soc. Am. **84**, 101–114 (1988)]. These show that the identification of nonsense syllables can be predicted as the product of the probabilities of identification of its phoneme parts, while word identification is systematically higher. Simulation studies are reported here involving a ''factored'' perceptual model that consists of a set of phoneme-likelihood estimators whose outputs are modulated by prior probabilities related to lexical status. This model can approximate the patterns observed in human perception. Simulations were also run with nonfactorable models, where syllables and words involve information about unique stimulus properties that cannot be predicted from their constituent phonemes. Consistent with a conjecture of Allen [J. Allen, IEEE Trans. Speech Audio Process. **2**, 567–577 (1994)], such syllable template models do not produce behavior compatible with human performance. [Work supported by SSHRC.]

4:20–4:45

Discussion

THURSDAY AFTERNOON, 4 DECEMBER 1997 FRIARS/PADRE ROOM, 1:30 TO 3:15 P.M.

Session 4pSP

Signal Processing in Acoustics: Sound Quality

Patricia Davies, Chair
School of Mechanical Engineering, Purdue University, Ray W. Herrick Laboratories, West Lafayette, Indiana 47907

Invited Papers

1:30

4pSP1. Artificial neural network processing of time-varying loudness data to model annoyance of noise. Peter C. Laux (Prince Corp., One Prince Ctr., Holland, MI 49423) and Patricia Davies (Purdue Univ., West Lafayette, IN 47907)

In this talk, a procedure is presented for the development of an artificial neural network model of human annoyance to noise. This model development is based on using available sound quality metrics (loudness, sharpness, fluctuation strength, and roughness) that were calculated using post-process temporal filtering and numerical formulation operations on measured time-varying critical band loudness data obtained from a Bruel and Kjaer loudness analyzer. The artificial neural network (ANN) was initially developed to model Zwicker's equation for unbiased annoyance (UBA). Subsequent training of the artificial neural network was done by using a series of 700 noise signals that were evaluated for their annoyance by a pool of normal-hearing young adult U.S. born subjects in a parameter estimation subjective test procedure. The noise stimuli were varied in seven different properties, which included: peak

loudness, modulation frequency, modulation envelope shape, modulation depth, frequency band of modulated noise, addition of continuous complex tone structures, and the relative level of complex tone set to the underlying noise. The response data were averaged across subjects and the UBA-ANN model was retrained using the average subjective responses. The results of this new ANN–annoyance model are compared to other metrics for sound measurement/rating.

2:00

4pSP2. A new aircraft interior noise simulator for psychoacoustic testing. Brenda M. Sullivan and Clemans A. Powell (Fluid Mech. and Acoust. Div., NASA Langley Res. Ctr., Hampton, VA 23681, b.m.sullivan@larc.nasa.gov)

NASA is conducting a research program in passenger response to aircraft interior noise to develop tools for use in design decisions for interior noise treatments. A new interior simulator has been built at NASA–Langley to be used in this program. The simulator is a shell fitted with interior trim and seats from 737/727 aircraft. It contains five listening stations, each having a pair of headphones for binaural signal presentation. Binaural recordings made in interiors of a number of aircraft were processed on a workstation into 50 sound stimuli. These were played back over the headphones and analyzed for repeatability within and between headphones. Initial results indicated that ranges averaged within headphones were 0.3 dB (A weighted), 0.3 phons and 0.7 dB (unweighted). When averaged between headphones, ranges were 1.0 dB (A weighted), 0.9 phons and 1.4 dB (unweighted). A first test in the simulator presented stimuli from propeller airplanes digitally modified to reduce the tonal components. Regressions between subjects' preference responses and measured metrics indicate that the subjective response correlates as well with the arithmetic mean of measurements from the left and right ears of the headphones as with the energy or pressure sums, and better than with measurements from the worse ear.

Contributed Papers

2:30

4pSP3. Perceived unpleasantness of natural sounds: Ratio-scale measurement and psychoacoustic analysis. Wolfgang Ellermeier, Markus Mader (Inst. für Psychol., Univ. Regensburg, D-93040 Regensburg, Germany), and Peter Daniel (Neutrik Cortex Instruments, Regensburg, Germany)

Paired comparisons of a heterogeneous set of ten natural sounds were collected from 60 listeners in order to determine: (1) if the sensation of unpleasantness is judged consistently across a wide range of acoustic stimuli; and (2) which sound features contribute to that sensation. The judgments conformed with the highly restrictive *BTL model* [R. D. Luce, *Individual Choice Behavior* (Wiley, New York, 1959)], thus justifying ratio-scale representation of perceived unpleasantness. The resulting scale values varied by a factor of 100 (diesel engine versus jackhammer). While they were not predicted by differences in A-weighted sound-pressure levels, a linear combination of the psychoacoustic attributes of loudness, roughness, and sharpness accounted for 98% of the variance in perceived unpleasantness.

2:45

4pSP4. Calibration of sound quality recording and playback systems. Poul Ladegaard (Bruel & Kjaer, Skodsborgvej 307, DK-2850 Naerum, Denmark)

In sound quality, a head and torso simulator (HATS) plays as important a role as the recording media. Used with playback of the signals through headphones, it allows the most accurate substitute for actually having the listener in the live situation. The human perception of sounds is very sensitive to changes in sound level, frequency response, and directional information. The value of a subjective sound quality evaluation is directly dependent on how well these parameters are maintained correctly. The paper describes the design methods and goals for a good HATS. Also the requirements for choice of headphones is discussed. In the complete recording/playback process, there is a need for some carefully selected frequency equalizations; these will be treated in some detail. Special attention will be given to the level calibration of the entire system. Here, the CIC check facility comes in as a very attractive and relevant tool. Finally, the summary will discuss the few limitations of the entire recording/playback system. Hints on how to overcome these will be given.

3:00

4pSP5. Voice selection for speech synthesis. Ann K. Syrdal, Alistair Conkie, Yannis Stylianou, Juergen Schroeter (AT&T Labs.—Res., 180 Park Ave., Florham Park, NJ 07932), Laurie F. Garrison, and Dawn L. Dutton (AT&T, Holmdel, NJ 07733)

A TTS voice quality experiment was conducted to select a speaker and to evaluate synthesis techniques. Small-scale TTS diphone inventories using six professional female speakers who were pre-selected in an audition were recorded. Two types of inventories were recorded for each speaker: a series of nonsense words and a series of English sentences. Using these 12 inventories, two synthesis methods were compared: PSOLA [Charpentier and Moulines, Eurospeech '89] and Harmonic Plus Noise (HNM) [Stylianou *et al.*, Eurospeech '97]. Synthetic prosody closely modeled naturally spoken versions of the target utterances. Three fully synthetic (TTS) and two hybrid (i.e., partly recorded from the human speaker and partly synthesized) sentences formed the experimental stimuli for subjective testing. For references, two MNRU versions of the naturally spoken sentences were used: (a) Q10 (resembling low-end commercial 16-kbps encoded speech) and (b) Q35 (resembling high-quality telephone speech). Forty-one subjects rated intelligibility [I], naturalness [N], and pleasantness [P] on five-point MOS scales. A total of 936 ratings were collected from each subject. Repeated measures of analyses of variance (ANOVAs) were performed on the data. There were significant main effects of speaker, synthesis method, and inventory, plus interactions. It was found that (1) the best speaker consistently outperformed the others on all three rating scales; for the optimal combination of parameters, TTS ratings ranged (across speakers) as follows: [I] 3.64–2.94, [N] 3.36–2.7, [P] 3.34–2.53. (2) HNM outperformed PSOLA (consistently 0.25 points higher for [I], [N], [P] scores), and (3) the diphone inventory extracted from sentences was preferred over that extracted from nonsense words (with a significantly smaller difference of 0.10 for HNM than 0.19 for PSOLA).

Session 4pUW

Underwater Acoustics: Signal Processing and Land-to-Sea Acoustics

Ahmad T. Abawi, Chair
Ocean and Atmospheric Science Division, NCCOSC RDTE Division, Code D881, San Diego, California 92152-5001

Contributed Papers

1:00

4pUW1. Matched Doppler processing with multiple receivers. Stephen J. Searle (CRC for Robust and Adaptive Systems, CSSIP and Univ. of Adelaide, Commun. Div., Knowledge Systems Bldg., P.O. Box 1500, Salisbury, SA 5108, Australia) and Douglas A. Gray (Univ. of Adelaide, Salisbury SA 5108, Australia)

Matched Doppler processing (MDP) is a new technique that estimates the range and speed of a near-field narrow-band sound source as it moves past an underwater acoustic sensor by coherently matching the received signal against replicas. Replica vectors are generated according to a simple propagation model, which accounts for the effect of the Doppler phenomena on the signal phase. The data to be matched are the vector of complex outputs of Fourier transforms of contiguous blocks of hydrophone data and the matching is by means of a cost or correlation function, which processes the signal phase coherently. The main contribution reported is the extension of MDP to an array of receivers for localization of a source's azimuth and heading as well as range and speed. Two methods are considered: The first method performs single receiver MDP on each sensor independently, then sums the cost values. The second method processes the signal phase coherently across space and also time. Extensive simulations compare the performance of both methods for an array of varying apertures and the number of receivers. Further simulation indicates that the second method affords sidelobe suppression, which improves performance when multiple signals are present. Application to tracking multiple sources is suggested and discussed.

1:15

4pUW2. On the use of signal autocorrelation matching in localization algorithms. Kevin B. Smith, Joachim Brune (Dept. of Phys., Naval Postgrad. School, Monterey, CA 93943), and Ching-Sang Chiu (Naval Postgrad. School, Monterey, CA 93943)

Many new ''robust'' localization algorithms rely on the insensitivity of a high-order processor. In this work, the robustness of a simple, Bartlett-type processor based on matching broadband signal autocorrelation functions is investigated. Specifically, the autocorrelation of the time-domain signal and the autocorrelation of the frequency-domain response will be matched. In order to highlight weaker aspects of these quantities, logarithms of these functions will also be used. Measures of robustness to be examined include the size of the localization footprint on the ambiguity surface and the peak-to-sidelobe levels in the presence of environmental mismatch. A full-wave PE model is used to produce broadband replicas. Model-generated synthetic signals with complete knowledge of the environment will provide base-line results. [This work is supported by the Naval Undersea Warfare Center, Code 2121.]

1:30

4pUW3. Channel-sensitive matched filter processor: Sensitivity and optimization study. Georgios Haralabus and Angela D'Amico (SACLANTCEN Undersea Res. Ctr., Viale S. Bartolomeo 400, 19138 La Spezia, Italy)

The detection performance of the channel-sensitive processor (CSP) [D. Alexandrou and G. Haralabus, SACLANTCEN Report No. SR-263 (1997)] has been tested in dense multipath conditions. It has been demonstrated that for a known propagation channel the CSP method outperforms the conventional matched filter technique. However, in an uncertain environment, the probability of detection of CSP degrades according to the degree of mismatch between the assumed and the actual channel characteristics. It has been found that the processor is more sensitive to geometric parameters (source range and depth) than to environmental parameters (sound velocity profile, sediment-subbottom interface, sediment depth). To overcome the performance degradation due to channel mismatch, the CSP method has been utilized in conjunction with two nonlinear optimization algorithms: the classical simulated annealing (SA) and the multilayer simulated annealing (MUSA) method. At the expense of processing time, it has been found that the optimization methods reduce the channel mismatch effect and considerably improve the detection performance of the CSP.

1:45

4pUW4. Inverse noise rejection. James H. Wilson (Neptune Sci., Inc., 4711 Viewridge Ave., Ste. 150, San Diego, CA 92123), James E. Donald, and Albert S. Nuttall (Naval Undersea Warfare Ctr., Newport, RI)

Inverse beamforming (IBF) has been developed over the last decade by the authors, and has proved highly successful in numerous at-sea tests. IBF is a unique signal processing method because the noise field properties are first determined by inverse techniques as a function of frequency, azimuth, magnitude, and the ''widths'' of these three dependent variables. The statistical properties of the noise field are determined, and a beamforming algorithm and six-dimensional tracker were developed to reject the noise and pass the signal (whose properties are well studied). IBF applies to deep water only, and this paper addresses the development of a similiar noise rejection algorithm in shallow water. The shallow-water problem is more complex due to the normal mode propagation characteristics at low frequencies, and the diversity of noise mechanisms in shallow water. Noise measurements are critically needed in shallow water to validate this approach.

2:00

4pUW5. Acoustic array analysis using multiple processors in parallel. Nils Paz and Ronald W. Townsen (CST Images, 3836 Browning St., San Diego, CA 92107, csti@ix.netcom.com)

Acoustic arrays are expensive to deploy and require a multidisciplinary team of scientists and engineers to develop, deploy, and monitor. This paper addresses one particular effort developed as a commercial software package, the Underwater Acoustic Analysis Tool (UWAA) by CST Images. UWAA is an acoustic array analysis tool for pre- and post-deployment modeling and analysis. Typically, the modeling of acoustic arrays and their deployment has been developed as separate and fractured modeling efforts. The array laydown modeling was separate from the acoustic beam formation and the propagation modeling. In addition to integrating all the components necessary to deploy an acoustic array, this effort describes the capability to compute the modeling in a parallel multithreaded environment using attached array processors. UWAA has integrated the modeling of the 3-D sensor positioning and sensor response, 3-D beam formation, 3-D acoustic modeling, and complex acoustic array interaction for beam formation using a high-performance multiprocessor system.

2:15

4pUW6. AWSUM EAP: An environmentally sensitive adaptive fluctuation-based processing algorithm that exploits amplitude and phase. Ronald A. Wagstaff and Jackson A. Mobbs (Naval Res. Lab., Stennis Space Center, MS 39529-5004, wagstaff@nrlssc.navy.mil)

Signals that propagate from submerged sources to deep receivers in the ocean generally interact less with the fluctuation generation mechanisms near the sea surface than do clutter signals and noise which originate near the surface. As a result, signals from submerged sources generally have smaller fluctuations than clutter signals and noise. The order dependence in the amplitudes of fluctuations has been used to devise an environmentally adaptive signal processing algorithm that provides preferential gains for signals having smaller fluctuation amplitudes than those of clutter and noise [R. A. Wagstaff and J. A. Mobbs, J. Acoust. Soc. Am. **101**, 3027 (1997)]. Gains include increases in signal-to-noise ratio, clutter suppression, and unalerted automatic detection. Similar gains can be achieved by exploiting the fluctuations in the phases of the acoustic pressures in the same manner that the order dependence was exploited. By replacing the order dependence with phase dependence, a similar algorithm has been devised that is sensitive to both amplitude and phase fluctuations and still adapts to the input data. The resulting algorithm, designated the AWSUM environmentally adaptive phase (EAP) is described, and results from measured data are presented. [Work supported by ONR and NRL.]

2:30

4pUW7. Array beamforming in long-range deep-water environments. Elena Yu. Gorodetskaya, Alexander I. Malekhanov, Alexander G. Sazontov, and Nadezhda K. Vdovicheva (Inst. of Appl. Phys., Russian Acad. of Sci., 46 Ul'yanov St., 603600 Nizhny Novgorod, Russia, almal@hydro.appl.sci-nnov.ru)

Realistic estimations of the deep-water acoustic coherence effects on the array gain for linear and quadratic beamformers, optimal ones included, were obtained for the North-West Pacific environments. An advanced technique was developed to calculate the signal coherence under the basic assumption that the long-range acoustic fluctuations are caused by internal waves in the summer channel or surface wind waves in the winter channel. Simulations were carried out both for horizontal and vertical arrays for the frequency of 250 Hz and ranges 500–1000 km. The following effects were shown to be the most essential points: (i) angular displacement and degradation of the plane-wave beampattern; (ii) large-array gain loss; (iii) coherence-induced gain ''gap'' between the optimal quadratic and linear beamformers; and (iv) gain dependence on the ambient modal noise. The optimal quadratic beamformer was shown to reduce the gain loss at a cost of increased processor complexity: The number of partial weight-sum channels is equal to the number of the largest signal eigenvalues. In some environments, a proper performance/complexity was realized using suboptimal beamformers which are, therefore, of particular interest for large-array applications. [Work supported by RFBR.]

2:45

4pUW8. Near-field polarization processing of ice fracturing events. Yuriy V. Dudko and Henrik Schmidt (Dept. of Ocean Eng., MIT, Cambridge, MA 02139)

As part of the ONR Sea-Ice Mechanics Initiative (SIMI) experiments in the spring of 1994, arrays of high-resolution seismic sensors were deployed in areas of high seismicity identified by real-time processing of acoustic emission events recorded by a wide, horizontal aperture hydrophone array. One of the areas most rich in ice events was located on a small ice floe ($\sim 100 \times 100$ m) 4 km Northeast of the main camp, where the seismic array consisting of five 3-component geophones in a 70-m aperture pentagon was deployed. One of the methods employed for analysis of the near-field data from this deployment was motion-product seismograms introduced by J. E. White [Geophysics **24**, 288–298 (1964)]. In this method each of the horizontal components of ice motion was multiplied by the vertical component, with or without phase shift, and after integration, the two resulting products identified a vector pointing to the source of the seismic waves. The ability of this method to separate different polarizations of the seismic waves and to determine the direction to the source was especially useful for geophone array data, because other analysis methods for such data occasionally failed due to the overlapping of waves generated by different ice events. Using the polarization processing method the development of ice fractures in the array near field was successfully tracked in the time and spatial domains. One result of the polarization analysis was that these fractures seemed to mostly generate vertically polarized shear (SV) waves. [Research supported by ONR.]

3:00–3:15 Break

3:15

4pUW9. Detection and tracking of land vehicle activity by offshore underwater acoustic arrays. Gerald L. D'Spain, Lewis P. Berger, William S. Hodgkiss, William A. Kuperman, LeRoy M. Dorman, and William A. Gaines (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0704)

Land-based vehicle activity can be tracked by underwater acoustic sensors located at least as far as 3.4 km offshore. This capability is demonstrated with data from two near-surf-zone experiments as part of the Marine Physical Laboratory's Adaptive Beach Monitoring program. In one case, four tracked vehicles spaced every 150 m traveled down the beach. Spectra estimated from a single hydrophone of a bottom line array 1.5 km offshore in 12-m water show a high-pass character compared to land geophone recordings because of the coupling characteristics between the land seismic and underwater acoustic fields. Adaptive plane-wave beamforming over the 30- to 70-Hz band shows, at times, beam-peak-to-background levels exceeding 35 dB. The acoustic tracks agree with those from visual logs and suggest that the four vehicles can be enumerated acoustically, particularly when combined with results from the 120- to 130-Hz band. In a second case, plane-wave beamforming results for two tracked vehicles traveling on the beach and recorded by a bottom array 3.4 km offshore in 20-m water show a beam-peak-to-background level approaching 20 dB. A videotape shows the evolution of 2-D wave-number spatial spectra during the vehicle transit. [Work supported by ONR, Code 32.]

3:30

4pUW10. Coupling of land-based signals into the nearshore underwater acoustic field. Gerald L. D'Spain, Lewis P. Berger, LeRoy M. Dorman, William S. Hodgkiss, and William A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0704)

Small, controlled land detonations on and near the beach recorded by land geophones and by offshore seismoacoustic sensors are used to characterize the coupling between land seismic and underwater acoustic fields. These data were collected in the Marine Physical Laboratory's Adaptive Beach Monitoring program. The predominant arrival on land is an elastic surface wave. Although an ocean bottom interface wave can be seen clearly in seismoacoustic recordings just outside the surf zone, it attenuates to nearly background noise levels by the time it reaches a bottom hydrophone array 1.5 km offshore in 12-m water. Travel-time and frequency/wave-number analyses indicate that the hydrophone array's received energy is composed mostly of dispersive body waves propagating as the lowest water-borne mode, with frequency-dependent phase and group velocities around 2.0 and 1.8 km/s, respectively. The frequency at which this lowest mode cuts off is the corner frequency of the high-pass filter representing the propagation from land to the underwater acoustic field. These results and numerical modeling, guided by insights obtained from classic work on acoustic propagation in a fluid wedge, are used to explain offshore underwater recordings of land vehicle activity. [Work supported by ONR, Code 32.]

3:45

4pUW11. Sound radiation from motion-related ice processes in the central Arctic. Catherine Stamoulis (Dept. of Civil and Environ. Eng., MIT, Cambridge, MA 02139) and Ira Dyer (MIT, Cambridge, MA 02139)

The radiation characteristics of individual acoustic events from the central Arctic SIMI experiment data were studied, in the frequency range 10–350 Hz, with the purpose to identify the event generating ice mechanisms. Four event subpopulations and consequently four ice mechanisms were distinguished. The first subpopulation consisted of events attributed to floe unloading, a postfracture process, which was consistently detected in the ambient noise data following intervals of high ice fracture activity. The event radiation patterns were independent of azimuth and were modeled by a stationary vertical dipole. The second subpopulation consisted of events attributed to tensile fracture, the radiation pattern of which was modeled by a weighted superposition of longitudinal octopoles, modified by a Doppler factor to account for source motion. The last two subpopulations consisted of events attributed to shear fracture. A lateral octopole and a weighted combination of lateral and longitudinal octopoles, both modified by a Doppler factor, were used to describe the radiation characteristics of events in these subpopulations, respectively. The adequacy of the latter model indicated that events in the fourth subpopulation had been induced by shear fractures, which propagated through the formation of arrays of tensile cracks in their tips and edges.

4:00

4pUW12. Statistics of very deep ocean noise. H. M. Walkinshaw (Box 72, Peapack, NJ 07977)

Sea noise was measured continually over a period of 1 year on experimental installations 40 miles south of Bermuda. One-hour stretches of data were recorded daily on four hydrophones, two on the seafloor and two suspended off bottom in 2420 fm s. Analysis of spectrum levels at six frequencies from 0.1 to 1.5 kHz included computation of means, standard deviations, distributions, and higher-order statistical moments. Spectral shape, levels, and variability closely resembled sea noise measured at shallower receivers throughout the western North Atlantic. The suspended very deep hydrophones were 2 to 3 dB noisier than those on the bottom; other statistical properties of both suspended and bottomed receivers were alike. The mean monthly noise migrated through an annual cycle of frequency-dependent change, from 7 dB at 0.1 kHz to 12 dB at 1.5 kHz. This seasonal trend progressed smoothly from winter high to midsummer low, though not at a constant rate. Monthly probability distributions changed form correspondingly, becoming skewed for months with a rapidly shifting mean and tightly peaked for seasons of relative climatic stability. [Work supported by NavElecs.]

4:15

4pUW13. Acoustic image analysis by color mapping. Frank A. Boyle (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

Acoustic returns from different types of structures can appear very similar, making acoustic image classification difficult. Current research is aimed toward developing novel methods of generating acoustic images that convey relevant information for target identification and classificatin. A color-mapping method [Boyle and Chotiros, J. Acoust. Soc. Am. **99**(4), 2553(A) (1996)], which maps an image's spectral content into color has been applied to several examples of broadband acoustic data, with apparent success. Structural features that produce similar echoes in amplitude and duration often emerge with clear differences in color according to their spectral signature. Possibilities for using color to convey acoustic properties other than spectral content are also being explored. One such method maps the phase of a signal into colors, while another is based on wavelet filtering of signals. The presentation will include descriptions of color mapping and related methods, and examples of their application to acoustic data.

4:30

4pUW14. Uncertainty in fisheries echosounder calibrations. David A. Demer (Southwest Fisheries Sci. Ctr., P.O. Box 271, La Jolla, CA 92038) and Michael A. Soule (Sea Fisheries Res. Inst., Capetown, South Africa)

Calibration of echosounders for fish stock assessment are commonly performed using the standard sphere method. A complete calibration includes characterization of system sensitivity versus detection angle for measuring target strengths (TS) of individual scatterers and the equivalent beam angle for measuring backscattering strengths of insonified volumes. To explore the accuracy of the method, on-axis TS measurements were made of three standard spheres (23-mm copper and 33.0- and 38.1-mm tungsten carbide), under controlled conditions at a frequency of 120 kHz (bandwidth = 1.2 kHz). The TS measurements, derived by both integrated and peak intensities, were compared to their theoretical counterparts. Also, system sensitivities and split-beam detection angles were characterized versus target bearings. To investigate measurement precision, a commercial echosounder (Simrad EK500) was used to measure the TS of the spheres over 15-h periods with constant pulse duration (0.3 ms) and water temperature (19 °C). To characterize the temperature dependence of the calibrations, measurements of system gain and transducer admittance were made versus water temperature (0–17 °C). Finally, bootstrap simulations were used to estimate the combined uncertainty in 120-kHz echosounder calibrations using optimal standard spheres and over ranges of temperatures typically encountered in fisheries surveys.

4:45

4pUW15. ATOC signal enhancement using adaptive filtering. Gary E. J. Bold and Sze M. Tan (Dept. of Phys., Univ. of Auckland, Private Bag 92019, Auckland, New Zealand)

A moored, autonomous recording system deployed off the east coast of New Zealand in early 1996 acquired signals from the Pioneer seamount ATOC source. Unfortunately, the data were corrupted by a very strong, aliased interfering signal electrically coupled from a malfunctioning power supply which drifted in frequency by over 400 Hz, and also by a mechanically coupled vibration at the rotation frequency of the hard disk used to store the data. Pulse compression of the biphase modulated psuedorandom sequence used to encode the transmission spreads the energy in these signals across the desired signal's spectral passband, degrading the signal-to-noise ratio of the receptions. Since both unwanted components are unstable in frequency, classical digital filters are unable to eliminate them. However, an adaptive LMS filter has been used to track and virtually remove the interfering signals before pulse compression, resulting in signal-to-noise ratio gains of up to 6 dB.

Meeting of Accredited Standards Committee S1 on Acoustics

J. P. Seiler, Chair S1
U.S. Department of Labor, Cochran Mill Road, P.O. Box 18233, Building 038, Pittsburgh, Pennsylvania 15236

G. S. K. Wong, Vice Chair S1
Institute for National Measurement Standards, National Research Council, Ottawa, Ontario K1A 0R6, Canada

P. D. Schomer, Chair, U. S. Technical Advisory Group (TAG) for ISO/TC 43, Acoustics
U. S. CERL, P.O. Box 9005, Champaign, Illinois 61826-9005

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, Building 233, Room A149, Gaithersburg, Maryland 20899

Accredited Standards Committee S1 on Acoustics. Working group chairs will report on their preparation of standards on methods of measurement and testing, and terminology, in physical acoustics, electroacoustics, sonics, ultrasonics, and underwater sound. Work in progress includes measurement of noise sources, noise dosimeters, integrating sound-level meters, and revision and extension of sound level meter specifications. 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 S3 on Bioacoustics

L. S. Finegold, Chair S3
USAF Armstrong Laboratory, Noise Effects Branch AL/OEBN, 2610 Seventh Street, Wright–Patterson Airforce Base, Ohio 43433-7901

R. F. Burkard, Vice Chair S3
Hearing Research Laboratory, State University of New York at Buffalo, 215 Parker Hall, Buffalo, New York 14214

P. D. Schomer, Chair, U. S. Technical Advisory Group (TAG) for ISO/TC 43, Acoustics
U. S. CERL, P.O. Box 9005, Champaign, Illinois 61826-9005

J. Erdreich, Chair, U. S. Technical Advisory Group (TAG) for ISO/TC 108/SC4, Human Exposure to Mechanical Vibration and Shock
Ostergaard Acoustical Associates, 100 Executive Drive, West Orange, New Jersey 07052

H. E. von Gierke, Vice Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 43, Acoustics and ISO/TC 108/SC4, Human Exposure to Mechanical Vibration and Shock
1325 Meadow Lane, Yellow Springs, Ohio 45387

V. Nedzelnitsky, U.S. Technical Adviser (TA) for IEC/TC29, Electroacoutics
National Institute of Standards and Technology, Building 233, Room A149, Gathersburg, Maryland 20899

Accredited Standards Committee S3 on Bioacoustics. The current status of standards under preparation will be discussed. In addition to those topics of interest, including hearing conservation, noise, dosimeters, hearing aids, etc., consideration will be given to new standards which might be needed over the next few years. Open discussion of committee reports is encouraged.

Scope of S3: Standards, spectifications, methods of measurement and test, and terminology in the fields of mechanical shock and physiological acoustics, including aspects of general acoustics, shock and vibrations which pertain to biological safety, tolerance, and comfort.

Session 5aAB

Animal Bioacoustics: Instrumentation in Animal Bioacoustics I

Mardi C. Hastings, Cochair
Department of Mechanical Engineering, Ohio State University, 206 West 18th Avenue, Columbus, Ohio 43210-1107

Charles R. Greene, Cochair
Greenridge Sciences, Inc., 4512 Via Huerto, Santa Barbara, California 93110

Chair's Introduction—9:00

Invited Papers

9:05

5aAB1. Phantom echo generation in dolphin echolocation experiments. Roland Aubauer and Whitlow W. L. Au (Hawaii Inst. of Marine Biol., Univ. of Hawaii, P.O. Box 1106, Kailua, HI 96734)

In behavioral experiments where real targets are used to investigate dolphin echolocation, it is often very difficult to extract the relevant echo parameters that the animals use to discriminate or classify. The complex relationship between the physical dimensions and the reflection characteristic of real targets prevents separate control of various echo parameters of the stimuli presented in an echolocation experiment. A new echo simulation method presented in this paper avoids this problem. Dolphin echolocation sounds are transformed with the target impulse response into artificial echoes, which are played back to the animal. The developed phantom echo system is implemented on a digital signal processing board and gives an experimenter fully programmable control over the echo generating process and the echo structure itself. Echoes of several underwater targets are simulated to evaluate the quality of the method. A comparison of simulated echoes with the original echoes demonstrated a very good agreement independent of the incident signal (cross-correlation coefficient >0.98). [Work supported by DAAD and DFG.]

9:25

5aAB2. Bioacoustics measurement system control and coordination using LABVIEW. James J. Finneran (Biosci. Div. Naval Command, Control and Ocean Surveillance Ctr., RDT&E Div., 49620 Beluga Rd., San Diego, CA 92152-6266) and Mardi C. Hastings (Ohio State Univ., Columbus, OH 43210)

Although LABVIEW is well known for its ability to generate custom applications with graphical user interfaces, its main strength lies in its compatibility with a wide variety of third-party hardware and software. LABVIEW instrument drivers exist for hundreds of devices from manufacturers such as Hewlett–Packard, Tektronix, Keithley, B&K, Wavetek, and Tucker–Davis. These instrument drivers and its support for OLE, DDE, and the GPIB protocol enable LABVIEW to control complex measurement systems containing a wide variety of devices. Data analysis is facilitated by built-in functions for digital filtering, signal processing, and many common statistical operations. This paper discusses some of the basic features of LABVIEW as well as a specific application: the use of LABVIEW to control an integrated testing environment developed to measure the acoustically-induced motions of the peripheral auditory organs in bony fishes using a noninvasive ultrasonic measurement system. In this application, a personal computer running LABVIEW was used to control and coordinate the measurement system, which included a DSP board, motorized positioning system, and video capture board. The PC also contained a GPIB interface, which was used to interface LABVIEW with a spectrum analyzer, lock-in amplifier, oscilloscope, digital multimeter, and function generator. [Work supported by ONR.]

Contributed Papers

9:45

5aAB3. A portable virtual instrument for collection of cetacean auditory evoked potentials. David A. Helweg, Donald A. Carder, and Sam H. Ridgway (NCCOSC RDT&E DIV., Code D351, 49650 Acoust. Rd., Rm. 108, San Diego, CA 92152-6254, helweg@manta.nosc.mil)

Currently, there are no direct measurements of the frequency-dependent hearing sensitivity of any threatened or endangered cetacean species. Auditory-evoked potentials (EP) allow rapid measurement of cetacean audiograms without the need for the training required to collect behavioral audiograms. A portable EP virtual instrument (VI) has been created that runs on a laptop computer using a commercial PCMCIA DSP card. The VI provides user control over stimulus design, stimulus projection schedules, and collection of evoked potentials via an intuitive graphic user interface. Audiograms can be generated for frequencies up to 8 kHz. EP audiometric methods and the portability of the EP VI makes this system an ideal component of a stranding toolkit for Stranded Whale Auditory Action teams.

10:00–10:15 Break

10:15

5aAB4. Active impedance control in a water-filled waveguide for low-frequency bioacoustic testing. James J. Finneran and Mardi C. Hastings (Dept. of Mech. Eng., Ohio State Univ., Columbus, OH 43210)

A challenging problem in studying low-frequency sound reception in aquatic animals is creating a suitable underwater acoustic testing environment. For studies of sound reception in fishes, a plane traveling wave is often a desirable stimulus because the relationship between acoustic pressure and particle velocity is well known. At low frequencies, the long

acoustic wavelengths in water rule out obtaining traveling waves in rectangular aquaria. In this study, a cylindrical, water-filled acoustic waveguide, with an active termination was used to generate constant frequency, plane traveling waves in the 12–400-Hz frequency range. The waveguide was constructed of acrylic tubing with an NRL J-13 transducer flanged to each end. One J-13 was the primary source and generated continuous harmonic waves. The active control system measured the transfer function between two hydrophones located inside the tube and used a pattern search algorithm to adjust the secondary source amplitude and phase in order to drive the measured transfer function to that for an unbounded wave. The active control system was able to reduce the terminating reflection coefficient to below 0.05 within the frequency range 12–400 Hz. [Work supported by ONR, Grant No. N00014-94-1-0337.]

10:30

5aAB5. An autonomous acoustic recorder for shallow arctic waters. Charles R. Greene, Jr. (Greeneridge Sci., Inc., Santa Barbara, CA 93110)

Autonomous acoustic recorders permit full-time environmental monitoring for extended periods. The device described here has been used on the ocean bottom in depths to 44 m. It uses a 4-Gbyte disk to record sounds sampled continuously at 1 kHz (500-Hz bandwidth) except that for 1 min every 14 m 24 s the sample rate is doubled. The lower sample rate is for monitoring the sounds of seismic surveys (airgun pulses) and bowhead whale calls. The higher sample rate is for ambient noise analysis. Capacity is adequate for 24 days. Retrieval is by grappling and has been used on 100 units that operated for 15 days in the Alaskan Beaufort Sea during September 1996. [Work supported by BP Exploration (Alaska), Inc.]

10:45

5aAB6. A microprocessor-based tag to record acoustics experienced and generated by migrating northern elephant seals. William C. Burgess (SRI Intl., 333 Ravenswood Ave., Menlo Park, CA 94025, burgess@sri.com), Peter L. Tyack (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), Burney J. LeBoeuf, Daniel P. Costa, and Daniel E. Crocker (Univ. of California, Santa Cruz, CA 95064)

Two microprocessor-based acoustic recording tags, deployed on northern elephant seals migrating from California to Alaska, were recovered in August 1997 after over 4 months at sea. The tags' hard disks contained measurements of pressure, temperature, and ambient noise, as well as acoustic signatures of swim speed, swim stroke rate, respiration, and cardiac function, obtained as one subject (''Trane'') swam across the northeastern Pacific and the other (''Rocky'') north along the West Coast. The subjects dove an average of 58 (Trane) and 81 (Rocky) dives per day during the 26 days that logger batteries supported data acquisition; maximum dive depths reached 780 m (Trane) and 770 m (Rocky). To function over prolonged periods under these demanding conditions, the tags were designed with robust enclosures and software- and hardware-based power conservation techniques. The hydrophones were potted in rear-facing conical resin housings, which were found empirically to reduce flow noise over 100–200 Hz by 10–15 dB compared with semicylindrical and tapered cylindrical hydrophone housings. The results suggest that hydrophone tagging methods offer a comprehensive and reliable means of sampling acoustic stimuli and associated behavior for free-ranging marine animals over long periods at sea. [Work supported by ONR and MBARI.]

11:00

5aAB7. An acoustic transponder based tracking, data-telemetry, and logging system for real-time monitoring of 3-D positions and collection of other data from large marine vertebrates. Marco Flagg (Desert Star Systems, 761 Neeson Rd., Ste. No. 9, Marina, CA 93933, mflagg@desertstar.com), Thomas F. Norris (Bio-waves, San Diego, CA 92123), and Donald Croll (Univ. of California, Santa Cruz, CA 95060)

An acoustic tracking/telemetry/data-logging system is being developed for real-time monitoring of underwater movements and behaviors of free-ranging whales from a vessel. This system consists of an acoustic transponder ''tag'' attached to the animal. Upon interrogation from a ''surface'' station, the transponder produces a high-frequency (38-kHz) signal. An array (either short or ultrashort base line) of transducers at the surface station receive and compute the underwater location of the tag. Locations are computed based on the return time and time or phase delay of the received signal. Multiple tags can be monitored simultaneously and relative underwater positions can be integrated with data from a GPS on the tracking vessel (resulting in a latitude, longitude, and depth). A microprocessor on the tag controls basic functions and can be programmed to record and store up to seven channels of data from attached sensors (e.g., water temperature, velocity, etc). Data are stored internally on the tag's RAM memory or can be telemetered acoustically to the surface station. Preliminary field testing in the open ocean resulted in an effective range of over 1 km. This system is being developed for determining movements and eventually acoustic behaviors of individual blue and fin whales. [Work supported by ONR.]

FRIDAY MORNING, 5 DECEMBER 1997 COUNCIL ROOM, 8:15 TO 11:15 A.M.

Session 5aED

Education in Acoustics: Informal Science Education in Acoustics

Thomas D. Rossing, Chair
Physics Department, Northern Illinois University, DeKalb, Illinois 60115

Invited Papers

8:15

5aED1. Learning about sound with interactive exhibits. Elsa Feher[a] (Ctr. for Res. in Mathematics and Sci. Education, San Diego State Univ., San Diego, CA 92182)

What and how can visitors learn from the interactive exhibits in a science museum? Using examples of actual exhibits that deal with sound, we will examine the process of developing them, and the ways in which visitors use them and interpret the phenomena that are displayed. [a]Former director, Reuben Fleet Sci. Ctr.

5a FRI. AM

8:45

5aED2. Acoustical considerations in developing the educational exhibition Mostly Music. David Taylor (Director of Sci. and Exhibits, Pacific Sci. Ctr., 200 Second Ave. N., Seattle, WA 98109)

Mostly Music is a new exhibition designed and developed by the Pacific Science Center. It will visit nine museums around North America during its 3-year tour. Developing an exhibition on the science and technology of music and musical instruments required consideration of varying background sound levels depending on the number of visitors in the exhibition, cross-talk between exhibits, requirements of sound fidelity and sound isolation, and sound reinforcement in some instances. Design case study will be presented along with anecdotal comments on planned versus actual results on individual exhibits.

9:15

5aED3. The Connections Project: An interactive look at the fine arts, mathematics, and physics. Christopher J. Chiaverina (New Trier High School, 385 Winnetka Ave., Winnetka, IL 60093)

Members of New Trier High School's fine arts, mathematics, and physics departments are using interactive exhibits to illustrate connections between seemingly disparate subjects areas. Teachers from these three departments, and their students, are producing hands-on displays that encourage learners of all ages to discover elements common to the arts, mathematics, and physics. This initiative has produced over 60 interactive exhibits that are being used in venues such as classrooms, instructional laboratories, school corridors, and art galleries. A number of exhibits demonstrate acoustical principles and their relationship to other disciplines. These and other Connections Project exhibits will be discussed.

9:45–10:00 Break

10:00

5aED4. Teaching with Exploratorium sound exhibits. Paul Doherty (Director of the Teacher Inst., Exploratorium, 3601 Lyon St., San Francisco, CA 94123)

The Exploratorium is a hands-on museum of science, art, and human perception. The museum includes many unique and educationally valuable exhibits dealing with the science of sound. Slides showing interesting exhibits will be presented, such as: the ''Sound Column,'' which demonstrates the sensitivity of the human ear to pressure fluctuations, not displacement fluctuations; the ''Air Reed,'' which uses Schleiren techniques to show the motion of air that drives a flute, and ''Find the Highest Note,'' an exhibit that not only shows Shepard tones but also allows visitors to test themselves with the tritone paradox. This presentation will discuss how these exhibits and more are used in our teacher education programs.

Contributed Paper

10:30

5aED5. A new graduate program in acoustics. David A. Brown and Karen Payton (Dept. of Elec. and Comp. Eng., Univ. of Massachusetts, N. Dartmouth, MA 02747, dbrown@umassd.edu)

The University of Massachusetts Dartmouth is offering a new graduate program in the area of Marine Acoustics leading to the degree of Masters of Science in Electrical Engineering with a specialization in Marine Acoustics. The Marine Acoustics specialization is multidisciplinary in nature bringing together the physical, mechanical, and electrical engineering aspects of the study of the production, propagation, and reception of underwater sound. The graduate program is currently running on campus as well as off-site at the Naval Undersea Warfare Center (NUWC) in Newport for as many as 60 students. Students in the NUWC program have completed the first 4 courses of an 11-course program consisting of a Mathematics Review (ECE403), the Fundamentals of Acoustics (ECE 557), Random Signals (ECE 521), Underwater Acoustics I (ECE 597), Underwater Acoustics II (ECE 597), Acoustic Transduction (ECE699), Digital Signal Processing (ECE574), Sonar Signal Processing (ECE575), Sonar Systems Engineering (ECE675), and a graduate thesis or two additional graduate courses (e.g., Computer Systems, Nonlinear Acoustics, etc.). The program is expected to provide the student with a good foundation in acoustics. Details will be elaborated on during the presentation. [Program sponsored by NUWC and UMass Dartmouth.]

Invited Paper

10:45

5aED6. Demonstrating the physics of brass musical instruments. Brian W. Holmes (Dept. of Phys., San Jose State Univ., San Jose, CA 95192-0106)

Brass musical instruments, unlike woodwinds, have complicated shapes: a cup-shaped mouthpiece, a conical leadpipe, a cylindrical valve section, and a flared bell. A trumpet will be built, allowing the acoustical significance of these parts to be demonstrated. It will be shown why brasses rely on valves to get different notes, unlike woodwinds, which use side holes. Before the invention of the valve, hornplayers learned to augment their meager supply of open notes by changing the position of the hand in the bell. Although the modern horn uses valves, the right hand remains important. The acoustical and musical significance of the hand will be explained and demonstrated. The talk concludes with a performance of Beethoven's Sonata for Horn and Piano, Op. 17, performed on a valveless horn similar to those in use in Beethoven's time.

Session 5aNSa

Noise and Engineering Acoustics: Measurement Errors, Instrumentation and Calibration

Robert A. Putnam, Chair
Westinghouse PGPG, MC504, 4400 Alafaya Trail, Orlando, Florida 32826-2399

Invited Paper

8:15

5aNSa1. Perspectives on uncertainty, tolerance, and precision in outdoor sound measurements. Robert A. Putnam (Westinghouse MC 590, 4400 Alafaya Trail, Orlando, FL 32826, putnam.r.a@wec.com)

Outdoor acoustical surveys typically involve greater variations in sound levels than indoor or laboratory measurements, whether the object is the determination of ambient statistical information, or the characterization of outdoor noise sources, such as for regulatory compliance purposes. Historical usage of each of the terms uncertainty, tolerance, and precision has been inconsistent and often vague, which leads to misunderstanding among all parties concerned with the definition of sound levels. The thing that has come to be referred to as uncertainty in outdoor sound measurements usually includes tolerance and precision. The field of acoustics is making progress in standardizing the definitions and uses of these, often misused, terms. This paper will discuss the perspectives of ISO, ANSI, ASME, and ASTM toward these terms, especially in light of recent standardization efforts. The use of the term tolerance, it will be recommended, should properly be a narrow one, as it is presently given in ASME performance test codes. This paper will summarize concise definitions and recommendations for standard usage, together with classification of the constituent elements of each term, in order to promote consistency within the acoustical community, thereby helping to insure improved clarity on the subject among all concerned parties.

Contributed Papers

8:45

5aNSa2. Extension of near-field acoustical holography to measurements in a moving medium. Richard J. Ruhala, David C. Swanson, and Courtney B. Burroughs (Grad. Program in Acoust., Penn State Univ., State College, PA 16804)

Near-field acoustical holography (NAH) has been widely used to study acoustic radiation from stationary sources. Recently, it has been applied to study tire noise on a vehicle in which the source and receiver are both moving in the medium. Since the source and the measurement plane are moving at the same speed, the frequency Doppler effect is absent. However, a wave-number Doppler effect exists. This leads to errors when reconstructing the acoustic field both toward and away from the source, even at low Mach number speeds. To investigate these errors, a point source is studied analytically using planar NAH with flow in one direction. The effect of the moving medium is noted by a shift in the radiation circle in the wave-number spectrum. From the nonlinear wave equation, a k-space Green's function and a k-space filter are developed that includes the effects of the moving medium. The newly developed methods are then applied to tire noise using measured data. [Work supported by DOT.]

9:00

5aNSa3. Calibrations of low level sound calibrators. Lixue Wu and George S. K. Wong (Acoust. Standards, Inst. for Natl. Measurement Standards, Natl. Res. Council, Ottawa, ON K1A 0R6, Canada)

Sound calibrators are designed to generate known sound-pressure levels at specified frequencies. Performances of calibrators are usually verified by accredited acoustical test laboratories. According to the latest standard revision [IEC 60942, Ed. 2, Sound Calibrator], the level of ambient noise reaching the microphone during testing of the calibrator shall be at least 30 dB lower than the measured sound-pressure level for classes 0 and 1 calibrators, and at least 25 dB lower than the measured sound-pressure level for class 2 calibrators. For calibrators, high-ambient noise will affect the uncertainty of the measurement, and this is particularly severe when the calibrator is operating at a relatively low sound-pressure level, such as 74 dB. In this paper, methods to reduce the effect of ambient noise on calibrators are presented. The uncertainties of the measurement associated with the methods are estimated. The measured sound-pressure levels based on the new methods are compared with those measured in an anechoic chamber using the traditional method.

9:15

5aNSa4. Analog and digital signal processing in a high-performance miniature sound-level meter. Jack Goldberg (JG Assoc., 8057 Vickers St., San Diego, CA 92111) and George S. K. Wong (Natl. Res. Council, Ottawa, ON K1A 0R6, Canada)

The design of a relatively low-cost Type 2 pocket-size sound-level meter with an embedded 8-bit microcontroller is described. The key design goal was to provide 110 dB of dynamic range without the need for user-controlled range switching. The miniature instrument meets all the current ANSI S1.4-1983 (R 1990) and IEC 651 (1979) sound-level meter standards. Relevant features in the future IEC 1672 sound level meter standard, such as level linearity, tone-burst response, over-range and under-range indication, measurement accuracy, and resolution are reviewed. Due to judicious separation of the functions performed by analog and digital hardware and software, the wide dynamic range is implemented with only 8 bits of A–D conversion. Extensive type testing confirmed that the instrument can satisfy the requirements of the standards over the required temperature range. [Work supported by Etymotic Research, Elk Grove Village, IL.]

Session 5aNSb

Noise: Subjective Aspects and Effects of Noise

Karl S. Pearsons, Chair
BBN Systems and Technologies, 21128 Vanowen Street, Canoga Park, California 91303

Contributed Papers

9:45

5aNSb1. Military aircraft noise and the outdoor recreationalist. Bartholomew Elias (Noise Effects Branch, Armstrong Lab., AL/OEBN, 2610 Seventh St., Wright–Patterson AFB, OH 45433-7901)

In response to growing concerns over aircraft noise impacts on outdoor recreationalists, the Department of Defense has initiated a collaborative research program with the National Park Service to study this issue. There are three main objectives to this project. The first objective is to assess park and wilderness land usage by outdoor recreationalists and utilization of military training routes and operations areas within park and wilderness boundaries. The second goal is to assess the response of outdoor recreationalists to military overflights. Recent Air Force sponsored research has examined park visitor reactions to military overflights along a short-hike interpretive trail. However, further research is needed to assess whether outdoor recreationalists seeking more remote back-country wilderness experiences are less tolerant of aircraft overflights. The final goal is to recommend, implement, and assess the effectiveness of various mitigation strategies. Promising mitigation strategies include remapping airspace and establishing operational procedures to minimize overflights of specific noise-sensitive park and wilderness areas, scheduling aircraft operations to minimize the temporal congruence of aircraft and outdoor recreationalists within a given area, and informing outdoor recreationalists of potential overflights within a given area while providing alternative recreational areas where aircraft noise is minimal.

10:00

5aNSb2. Measurements of personal aircraft noise exposure of outdoor recreationists. Matthew Sneddon, Sanford Fidell, and Karl Pearsons (BBN Systems and Technol., 21128 Vanowen St., Canoga Park, CA 91303)

The at-ear noise exposure of hikers created by sporadic aircraft overflights in an outdoor recreational setting was measured on public lands underlying an active military operations area. Cumulative noise levels for entire hikes and for shorter time intervals were compared with conventional (place-oriented) noise monitoring records. Variability in personal aircraft noise exposure, even of hikers on the same itinerary but at different times of day, was so great as to call into question the adequacy of conventional noise exposure measurement and aircraft noise exposure modeling as representations of the actual aircraft noise experience of outdoor recreationists. Distributions of closest points of approach of aircraft to hikers (derived from radar flight track information and time-stamped geo-coding of hikers' paths) were examined as an explanatory mechanism for the variability. [Work supported by U.S. Air Force Air Combat Command.]

10:15

5aNSb3. Methodological study of sleep motility measurements. Laura Silvati and Karl S. Pearsons (BBN Systems and Technol., 21128 Vanowen St., Canoga Park, CA 91303, lsilvati@bbn.com)

A methodological study was undertaken to investigate the sensitivity of sleep motility to noise-induced disturbances. Indoor noise exposure and sleep motility of 22 test subjects were measured for 109 subject nights under a range of conditions of noise exposure and familiarity of sleeping quarters. Recording accelerometers stored the number of times that vertical motion of the wrist-worn devices exceeded 0.0002 g at frequencies below about 3 Hz during a 2-s interval. Noise monitors continuously recorded 2-s time histories of indoor noise levels throughout each night of data collection. Additional sequences of varying epoch durations were constructed analytically from the 2-s intervals of noise levels and sleep motility data. Analyses of accelerometric recordings in 2-, 4-, 6-, 8-, 10-, 20-, and 30-s epochs included an examination of the relationship between synchronous indoor noise level and motility, description of motility observations on a nightly basis, characterization of general serial dependence in motility measurements, and characterization of motility measurements in consecutive epochs with and without movement. Characterization of sleep motility in epochs of short duration provided limited practical benefits for measurement of noise-induced sleep disturbance.

10:30

5aNSb4. Nighttime sleep disturbance in residential communities—A review of recent U.S. field research and a recommended new dose–response relationship. Lawrence S. Finegold (USAF Armstrong Lab., Noise Effects Branch, 2610 Seventh St., 45433 Wright–Patterson AFB, OH 45433, lfinegold@osprey.al.wpafb.af.mil)

This paper provides a brief review of field research on nighttime sleep disturbance in residential communities which has been conducted in the United States over nearly the past decade. It compares the research methodology for studying sleep disturbance that has evolved during this series of research projects with the research approaches predominantly used in Europe. The focus is on transportation noise sources, primarily aircraft noise. It also offers a new dose–response relationship based primarily on the U.S. field data collected around civilian airports and military airbases. The new sleep disturbance prediction curve is recommended as an update to the interim curve adopted by the Federal Interagency Committee on Noise (FICON) in 1992. At that time, there was a lack of consensus concerning whether both laboratory and field research data should be used in the development of a new prediction curve. Thus the FICON curve combined the major field and laboratory data published in the U.S. up to that time in an admittedly interim curve. The currently proposed new prediction curve uses only data from field studies and incorporates some of the European research data.

10:45

5aNSb5. Comparison of noise metrics for predicting the annoyance of aircraft overflight noise. Karl Pearsons, Richard Howe, Matthew Sneddon, Laura Silvati, and Sanford Fidell (BBN Systems and Technol., 21128 Vanowen St., Canoga Park, CA 91303)

Two groups of 30 audiometrically screened test participants judged the relative annoyance of two comparison (variable level) signals and 30–34 standard (fixed level) signals in an adaptive paired comparison psychoacoustic study. The signal ensemble included primarily stage II and stage III aircraft overflights, as well as synthesized aircraft noise signatures. Test signals were presented for judgment as heard indoors (test 1) and outdoors (test 2), in the presence of continuous background noise, under free-field listening conditions in an anechoic chamber. For both tests, analyses of the performance of 30 noise metrics as predictors of these annoyance judgments confirmed that the more complex metrics were generally more ac-

curate and precise predictors than the simpler methods. EPNL was slightly less accurate and precise as a predictor of the annoyance judgments than a duration-adjusted variant of Zwicker's loudness level. [Research was supported by NASA, Langley Research Center, Hampton, Virginia.]

11:00

5aNSb6. Investigation of low frequency aircraft noise near airports. Stephen Lind and Karl Pearsons (BBN Systems and Technol., 21128 Vanowen St., Canoga Park, CA 91303)

Noise levels due to aircraft departures were measured at four locations near Minneapolis–Saint Paul International Airport and at seven locations near Los Angeles International Airport. A-weighted overall levels and one-third octave band levels between 25 and 80 Hz were obtained. Levels for 46 aircraft were analyzed for Minneapolis and 122 for Los Angeles. A linear regression was performed on the low-frequency levels versus the A-weighted levels for the locations to the sidelines of the active runways. The relationship between low-frequency noise and A-weighted sound pressure level for different distances was applied to A-weighted levels obtained from INM to estimate the low-frequency noise levels from a proposed runway. Estimates were made regarding which parts of the community would experience perceptible vibrations.

11:15

5aNSb7. Airbag noise hazard examined with a mathematical model of the human ear. G. Richard Price (U.S. Army Res. Lab, HRED, APG, MD 21005-5425, dprice@arl.mil)

When cat ears were exposed to an airbag impulse in an open, closed, or sealed passenger compartment, a mean-threshold shift of 60 dB occurred immediately and resolved into a 37-dB permanent threshold shift [Price *et al.*, J. Acoust. Soc. Am. **99**, 2464 (1996)]. A mathematical model of auditory hazard for the cat ear ''explained'' the results [Price and Kalb, J. Acoust Soc. Am. **99**, 2464 (1996)]; but there was still a question as to what the implications would be for human ears exposed to such impulses. A parallel model has been created for the human ear and the hazard re-evaluated. For the human ear, the presence of middle ear muscle activity is calculated to be a key element in preserving hearing as is the (paradoxically) protective effect of a closed or sealed passenger compartment as compared with one that is open. The individual exposed in a vented passenger compartment and who did not see the accident coming is, therefore, predicted to be at greatest risk. Algorithms are being developed and validated for connecting the model's output (auditory damage units) to projected changes in threshold for a human population whose susceptibility is normally distributed.

11:30

5aNSb8. Variation in estimates of excess risk among screened and unscreened populations. Mary M. Prince (Natl. Inst. for Occupational Safety and Health (NIOSH), 4676 Columbia Pkwy., R-16, Cincinati, OH 45226)

Analyses of data from the NIOSH Occupational Noise and Hearing Survey (OHNS) have previously focused on 1172 highly ''screened'' workers. This paper examines excess risk estimates for an additional 894 male workers (609 noise exposed and 285 controls) who were excluded for various reasons (i.e., nonoccupational noise exposure, otologic or medical conditions affecting hearing, prior occupational noise exposure). Data are analyzed by age, duration of exposure, and sound level (8-h TWA) for different definitions of noise-induced hearing impairment. Hearing impairment was defined as a binaural average hearing threshold level greater than 25 dB at selected frequencies (1–4 kHz, 1–3 kHz, and 0.5, 1, and 2 kHz). The analysis suggests that excess risk estimates are sensitive to population characteristics as well as definition of impairment. Lifetime excess risk estimates for noise-induced hearing impairment are similar for the unscreened and screened populations for 8-h TWAs below 85 dB. However, compared to the unscreened population, higher excess risk estimates are observed for the screened population at A-weighted levels greater than 85 dB. This difference in excess risk was most marked for hearing-impairment definitions that include 3 and 4 kHz.

11:45

5aNSb9. Noise control under successive fulfillment of jobs specified by various degrees of labor tension. Mark V. Levin (Dept. of Phys. Factors, Inst. for Occupational Health, 75 Saksagansky St., 252033 Kiev, Ukraine) and Alexander A. Menshov (Burlingame, CA 94010)

The Standards of occupational noise (A-weighted) in Russia, the Ukraine, and other countries in the former USSR are differentiated by degrees of labor tension (40–80 dB). Frequently there is a real necessity to solve the problem of noise control, which is summed up for a full-time day during which labor is exposed to different noise levels. The problem is addressed by the definition of equivalent sound shift level—the noise working loudness. This definition correctly sets forth one noise standard during a shift of one action, i.e., the fulfillment for a particular period by a worker of identical labor tension. The decision of this task is complicated by the presence during the shift of work distinguishing not only the noise levels but different labor tension. Under the circumstances, it is advisable to sum up noise control according to the quotient of noise loudness (QNL). It is defined by the overage in time exceeding the standards of sound level for every separate period of labor with different degrees of labor tension. The quotient of noise loudness (QNL) is calculated by the proposed formula. It allows one to make noise control under successive fulfillment of jobs specified by various degrees of labor tension.

Session 5aSC

Speech Communication: Prosody, Timing and Source Characteristics (Poster Session)

Jody Kreiman, Chair
Head and Neck Surgery, University of California, Los Angeles, 31-24 Rehab, Los Angeles, California 90024-1794

All posters will be on display from 9:00 a.m. to 12:30 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:00 a.m. to 10:45 a.m. and contributors of even-numbered papers will be at their posters from 10:45 a.m. to 12:30 p.m. To allow for extended viewing time, authors are requested to leave their posters on display until 5:00 p.m.

Contributed Papers

5aSC1. Comparison of pitch reset characteristics between children's filler syllable and starter syllable productions. P. J. Seymour, Shari R. Speer, Marc E. Fey, and Diane Frome Loeb (Dept. of Speech-Lang.-Hearing, Univ. of Kansas, Lawrence and Kansas City, 3910 Rainbow Blvd., Kansas City, KS 66160)

Children's filler syllables may assist in maintaining rhythmic envelopes and intonational phrasing, and may be markers of potential or actual morphosyntactic information. Therefore, they should exhibit pitch characteristics indicating coherence with other constituents in the same intonational phrase. They contrast with starter syllables, which serve more pragmatic functions such as discourse place holders. An acoustical analysis of pitch reset is presented which distinguishes between filler and starter productions. Pitch reset was evaluated through two measures: the difference in intensity (decibels) between syllable initiation and syllable termination, and boundary duration (milliseconds) from the offset of the targeted syllable to the onset of the next syllable or constituent. Results demonstrated that fillers and starters could be differentiated using pitch reset measures. Typically, fillers did not exhibit pitch reset, described here as a boundary duration less than 200 ms and intensity difference scores ranging from -15 to $+17$ dB. Starters, however, did typically exhibit pitch reset, described here as a boundary duration greater than 200 ms and intensity difference scores ranging from 0 to $+17$ dB. This pilot investigation suggests that fillers exhibit similar pitch reset acoustical characteristics to other syntactic constituents in the intonational phrase in which they are located. [Work supported by NIH.]

5aSC2. Interaction between prosody and discourse structure in a simulated man–machine dialogue. Robert Eklund (Telia Res. AB, Spoken Lang. Processing, S-136 80 Haninge, Sweden)

Automatic speech understanding systems are beginning to attain a level of sophistication where commercial applications are within reach. However, if humans and machines are ever going to communicate in a natural way, it is of vital importance that language modeling go beyond the sentence level. A profound understanding of discourse structure is required, and to this end, knowledge concerning how prosody interacts with other linguistic phenomena is needed. Not only will better prosodic modeling of discourse lead to better speech recognition/understanding, it will also yield more natural-sounding speech synthesis. This paper reports on a dialogue/prosody project at Telia Research, Sweden. A Wizard-of-Oz simulation of a computerized reservation system was used to collect realistic speech data. Fifty subjects were given three tasks each that entailed the reservation of flights, trains, car hire, and hotel reservations. To avoid linguistic influence on the subjects' utterances, the tasks were given as maps and icons. A ToBI-style analysis was applied, adapted to meet language-specific requirements. The dialogues were analyzed with regard to phrase boundaries, tones, disfluencies, syntax (functions/categories), new versus given information, and pitch range. This paper describes our observations concerning the interaction between prosodic, syntactic, and higher-level linguistic phenomena, such as discourse structure.

5aSC3. Categorical judgments of pathological voice quality. 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, jkreiman@ucla.edu)

Categorical judgments of vocal quality are often applied in studies seeking to synthesize particular voice types, and in studies of the acoustic correlates of particular qualities. However, listeners' ability to reliably classify voices as ''breathy'' or ''rough'' has never been demonstrated. In this study, 15 listeners decided whether each of 160 voice samples was or was not primarily rough in quality; in a separate experiment, a second group of listeners judged whether each voice was primarily breathy. Although listener agreement exceeded chance levels, only a few voices were consistently categorized as breathy or rough. Many voices with a high probability of being judged ''breathy'' also had a high probability of being judged ''rough.'' Finally, the likelihood of a ''rough'' or ''breathy'' response correlated highly with the rated severity of vocal pathology. These results confirm that breathiness and roughness are not independent concepts, and that neither captures information apart from overall severity of pathology. Thus categorical descriptions of a pathological voice are probably not useful clinical or research tools. [Work supported by NIH.]

5aSC4. Early use of pitch accent in Japanese spoken-word recognition. Takashi Otake (Faculty of Foreign Lang., Dokkyo Univ., 1-1 Gakuen-machi, Soka-shi, Saitama, 340 Japan) and Anne Cutler (MPI for Psycholinguist., Nijmegen, The Netherlands)

A gating experiment addressed the question of how early in the process of recognizing spoken Japanese words pitch–accent information may be exploited. Twenty-four pairs of Japanese words such as *nimotsu/nimono*, beginning with the same bimoraic CVCV sequence but with the accent pattern of this initial CVCV being HL in one word and LH in the other, were presented, in increasingly large fragments, to 36 native speakers of Tokyo Japanese. After presentation of each fragment, which was incremented in each case by one phoneme transition from the previous fragment, listeners recorded a guess regarding the word's identity and a confidence rating for that guess. The results showed that the accent pat-

terns of the word guesses corresponded to the accent patterns of the actually spoken words with a probability significantly above chance from the second fragment onwards, i.e., from the middle of the vowel in the first mora of the word. Accent correspondence averaged 79.6% at this point, rising to 89% by the fourth fragment (vowel of second mora). This demonstrates that Japanese listeners can exploit pitch–accent information effectively at an early stage in the presentation of a word, and use it to constrain selection of lexical candidates.

5aSC5. Simultaneous communication in beginning signers, Part I: Temporal characteristics of speech. Robert L. Whitehead (Appl. Lang. and Cognit. Res., Natl. Tech. Inst. for the Deaf, 52 Lomb Memorial Dr., Rochester, NY 14623-5604), Nicholas Schiavetti, Dale Evan Metz, and Thomas Farrell (State Univ. of New York, Geneseo, NY 14454)

Simultaneous communication combines spoken English with manual representations of English words by signs and fingerspelling. The purpose of this study was to investigate temporal characteristics of speech produced during simultaneous communication (SC) by beginning signers. Audio recordings of stimulus words embedded in sentences produced with speech only and SC were made by 12 students during the first and last weeks of a 15 week introductory sign language course. Temporal measures included: (a) sentence duration; (b) experimental word duration; (c) vowel duration of the experimental word; (d) interword interval before the experimental word; and (e) interword interval after the experimental word. Results indicated significant temporal differences between speech-only and SC conditions at both the beginning and the end of the sign language course. In addition, a pattern emerged in which beginning signers first signed between words in SC, thereby extending interword intervals, and then, at the end of the sign language course, shifted toward attempting to simultaneously sign while speaking the words, thereby elongating segmental temporal characteristics such as vowel duration. These results are consistent with previous findings regarding the effect of simultaneous communication on temporal characteristics of speech with experienced signers.

5aSC6. Pathologic discrimination through acoustic analysis of voice. Marcelo de Oliveira Rosa, Jose Carlos Pereira (Escola de Engenharia de Sao Carlos, USP, R. Dr. Carlos Botelho, 1465, C. P. 359, Sao Carlos, SP, Brazil, 13560-250, marcelol@lim.sel.eesc.sc.usp.br), Simone Adad Araujo, and Marcos Grellet (Faculdade de Medicina de Ribeirao Preto, Ribeirao Preto, SP, Brazil, 14049-900)

The acoustic analysis of voice from dysphonic patients has been studied, mainly for classification of voice quality in terms of roughness, breathiness, and hoarseness. By a data reduction of speech signal, researchers have produced several acoustic measures in order to relate them with a perceptual diagnosis. In this work, a proposal was made to modify acoustic parameters in order to support a higher sample quantity, instead of using only a data window by digitalized speech. Also, robust statistics is used to obtain realiable measures that estimate the abnormal condition of a voice signal as well as to discriminate several pathologics of the larynx and vocal cords, without any perceptual consideration. Using only seven measures from adaptative inverse filtering (Kalman and Wiener filters) and a neural network for pattern classification on 228 voice signals of speakers (20 groups of distinct types of pathologies and one of normal speakers) evaluated through videolaringoscopy, results of the pathologic discrimination showed 22% of uncertainty. Redrawing the patient sample in seven distinct groups (five with defined pathologies, one with all others pathologies, and one control), then the error in the pathologic discrimination is 18%. [Work supported by Capes.]

5aSC7. Acoustic cues to syntatic structure. Kris Tjaden, Lewis P. Shapiro (Dept. of Commun. Disord., San Diego State Univ., 5500 Campanile Dr., San Diego, CA 92182-1518), Janet Nicol (Univ. of Arizona, Tucson, AZ 85721), and David Swinney (Univ. of California at San Diego, La Jolla, CA 92093-0109)

Duration and fundamental frequency ($f0$) are used by listeners in on-line sentence processing as cues to syntactic structure [H. Nagel, L. P. Shapiro, and R. Nawy, J. Psycholing. Res. **23**, 473–485 (1994)]. Research suggests, however, that speakers may not reliably use duration and $f0$ to disambiguate syntactic structures [D. W. Albritton, G. McKoon, and R. Ratcliff, J. Exp. Psych. **22**, 714–735 (1996)]. The present study explored prosodic differences in a variety of syntactic structures for sentences read by five speakers. Speakers failed to utilize $f0$ or duration to disambiguate verbs followed by either a direct object or sentential clause; this finding does not support Nagel *et al.*'s (1994) results for one speaker. In contrast, intransitive verbs were characterized by longer durations and greater $f0$ excursion as compared to transitive verbs. Acoustic differences for noun phrases (NPs) following transitive and intransitive verbs also were present; NPs following transitive verbs were longer in duration and were associated with a steeper $f0$ change, as compared to NPs following intransitive verbs. The results suggest that speakers may distinguish some but not all syntactic structures using the prosodic cues of $f0$ and duration; implications for theories of on-line sentence processing will be discussed. [Work supported by NIH DC00494, DC03347.]

5aSC8. Rethinking the basics of declination. Yi Xu (2299 N. Campus Dr., Northwestern Univ., Evanston, IL 60208)

Declination, the tendency of fundamental frequency to gradually fall over the course of an utterance, has for a long time been treated as if it were a universal principle of speech intonation. Most intonation models adopt declination as a baseline upon which more local $F0$ patterns reside. A reexamination of the literature related to declination finds, however, that the existence of declination as an independent $F0$-determining mechanism is far from being firmly established. Instead, some recent as well as earlier studies are found to provide evidence that various linguistic factors may be responsible for generating different local as well as global patterns of intonation, which, taken as a whole, may sometimes give the appearance of declination. These factors include sentence focus, topic initiation, word stress, and lexical tone. When these factors happen to be neutralized, virtually flat $F0$ contours may occur, as have been observed in some tone languages. Furthermore, the commonly assumed declination rate is found to well exceed the possible rates of pitch decline due to global physiological limitations. It is concluded that, rather than being a basic principle of intonation, declination is more likely to be an artifact of various linguistic factors plus certain local physiological constraints.

5aSC9. Interactions among *F0*, duration, and amplitude in the perception of focus. H. Timothy Bunnell, Steven R. Hoskins, and Debra Yarrington (duPont Hosp. for Children/Univ. of Delaware, 1600 Rockland Rd., Wilmington, DE 19803)

$F0$, amplitude, and durational cues are considered the primary acoustic correlates of focus or sentence-level stress. However, questions remain regarding: (a) the degree to which each of these cues are necessary or sufficient for signaling focus; (b) the relative importance of each of these cues; and (c) how they interact in the perception of focus. Natural productions of the sentence ''Bob bought Bogg's box,'' in which focus was varied over each of the four words of the sentence were altered to produce prosodic cue neutralized versions. The alterations were applied singly and in all possible combinations to form eight experimental versions of each original sentence (the original and seven cue-neutralized versions). The sentences were presented to listeners with the task of identifying the focused item in the sentence. Results indicated that: (a) neutralizing any acoustic cue produced some degradation in performance, but even with all cues nullified, performance remained above chance for at least some words; (b) overall, $F0$ was more important than either amplitude or dura-

tion in signaling focus; and (c) $F0$ was more important early in the sentence while amplitude and duration played relatively more prominent roles late in the sentence. [Work supported by NIDRR and Nemours.]

5aSC10. Glottal airflow voice source characteristics of 11-year-old singers. Peta White (Roehampton Inst. London, Downshire House, Roehampton Lane, London SW15 4HT, UK)

The voice source of 18 girls and 10 boys from a music school was analyzed by inverse filtering the flow signal during sustained phonation of the vowel /a/ at three pitches ($F0$ between 220 and 330 Hz) using soft, mid, and loud voices. Subglottal pressure was measured as the oral pressure during the /p/ occlusion preceding the vowel. As compared to previous data on adult voices, the pressure values were high, and boys tended to use slightly higher pressures than girls. The influence of pitch and loudness variation was similar to that previously reported for adult voices. The ac flow amplitude was substantially smaller than previously reported for adults and was significantly higher in the boys than in the girls, while no significant sex differences were found for the children's closed quotient and for the level difference between the two lowest spectrum partials. Contrary to findings for adults, the dc flow during the quasiclosed phase was greater in the boys. Overall, the present findings were similar to those previously reported for untrained children's voices. Thus children's voice source characteristics can be more conveniently derived from singer subjects with their greater skill in independent control of pitch and loudness.

5aSC11. Combined simulation of two-dimensional airflow and vocal-fold vibration. Fariborz Alipour and Ingo Titze (Dept. of Speech Pathol. and Audiol., The Univ. of Iowa, Iowa City, IA 52242)

Self-sustained oscillation of the vocal folds was simulated by combining two-dimensional laryngeal airflow with multidimensional tissue movement. A finite-element model was used for the solution of viscoelastic waves in the tissue and a finite-volume method was used in the solution of Navier–Stokes equations for the airflow. A so-called ''shadow method'' simulated the glottal constriction in the flow model to avoid the complexity of grid movement. The two-dimensional flow equations were solved in an iterative manner until the given transglottal pressure was approximated. The flow solution was then used in the estimation of the aerodynamic forces on the tissue, required in the finite-element solution of tissue movement. The results indicate that inlet velocity profiles to the glottis are almost parabolic at any instant of time. The glottal velocity increases to its maximum at the glottal exit at the center of a jet, with values exceeding 40 m/s for 0.8 kPa of lung pressure. The jet velocity waveform is similar to that of an excised larynx and the pressure profiles are similar to those of steady flows in physical models. Also, the displacement of the inferior portion of the vocal folds leads the superior position in phase.

5aSC12. Native language determines the parsing of nonlinguistic rhythmic stimuli. Kiyomi Kusumoto and Elliott Moreton (Dept. of Linguist., Univ. of Massachusetts, Amherst, MA 01003)

Given a rhythmic sound stimulus, the hearer automatically parses it into recurrent units. Researchers since Woodrow [Arch. Psych. **14**, 1–66 (1909)] have consistently found that stimulus elements differing in duration are parsed iambically, while those differing in intensity are parsed trochaically. In an extensive typological survey, Hayes [*Metrical Stress Theory* (1995)] observes that no language has quantity-insensitive iambic feet, and suggests that the nonlinguistic parsing tendencies may be at the root of this language universal (the ''iambic-trochaic law''). However, the relevant nonlinguistic parsing experiments were done with native English speakers. We compared parsing behavior in native speakers of English and in two dialects of Japanese. Our findings suggest that native language (specifically, the structure of the minimal intonational phrase) determines parsing preference, not the other way around. The lack of quantity-insensitive iambs must have another source.

5aSC13. Resyllabification evidence for diphthongal off-glides as coda consonants. Bryan Gick (Haskins Labs., 270 Crown St., New Haven, CT 06510 and Dept. of Linguist., Yale Univ., New Haven, CT, bryan.gick@yale.edu)

The phonological status of diphthongal off-glides has long been controversial, with some researchers classifying them as consonants, others as vowels, and still others as vocalic features. The present experiment used electromagnetometer and video data to measure the effect of following context on the labiovelar off-glide of the diphthong /aw/ for three speakers of different dialects of American English. Kinematics of lip and tongue movement were measured, with findings showing the magnitude of lip aperture to be significantly smaller when followed by /a/ than when followed by /ha/, resulting in a constriction intermediate in size between that of the preconsonantal final off-glide and that of the onset glide /w/. Such sensitivity to following context indicates a partial resyllabification of the final off-glide into the onset of the following syllable, a property heretofore observed only in coda consonants. These results support the view that diphthongal off-glides are best analyzed as consonants occupying syllable coda position. [Work supported by NSF and by NIH Grant Nos. HD-01994, DC-02717.]

5aSC14. A five-year longitudinal investigation of duration and temporal variability in children's speech production. Bruce L. Smith and Mary Kay Kenney (Dept. of Commun. Sci. and Disord., Northwestern Univ., 2299 North Campus Dr., Evanston, IL 60208-3570)

Although cross-sectional (group) investigations of children's speech production have provided important findings about average performance and general patterns of development, one limitation of such studies is that they tend to obscure details concerning the development of *individual* children. In the present study, two children were recorded at approximately 8-month intervals for about 5 years as they produced 20 repetitions (per session) of several different target stimuli. One of the children was about $1\frac{1}{2}$ years old at the onset of the study, and the other was about 4 years of age. Acoustic measurements of various temporal aspects of their speech were obtained from spectrographic displays. The findings for both children showed certain patterns that supported the general claim (based on cross-sectional data) that segment and word duration and their associated temporal variability tend to decrease with increased age. However, there were also a number of instances in which the children's temporal variability and, in particular, their duration measures showed little or no change across the five years they were followed. Although the findings are from just two children, the implications of these longitudinal data are important for gaining a more detailed and comprehensive understanding of children's speech production development.

5aSC15. Pitch discrimination during breathy versus modal phonation (final results). Daniel Silverman (Dept. of Linguist., 4th Fl. FLB, Univ. of Illinois, Urbana, IL 61801)

This study completes an investigatation of pitch discrimination during breathy phonation versus modal phonation by manipulating acoustic parameters. Both breathy vowels and modal vowels from Jalapa Mazatec were digitized. Intensity was equalized, and $F0$ was manipulated by approximately 3-Hz increments (up to 12 Hz) employing the SOUNDEDITPRO16 sound editing program; all other acoustic parameters were held constant. Ten subjects each listened to 1000 pairs of these stimuli and judged whether they were the same or different. Subjects performed more accurately on modal vowel pairs than on breathy vowel pairs. Moreover, at the 3- and 6-Hz intervals, the difference between performances on modal pairs versus breathy pairs was significantly greater than the difference between modal pairs versus breathy pairs at the 9- and 12-Hz intervals. This indicates that, although frequency distinctions increase linearly in both phonation groups, a nonlinear discriminability relationship exists within each group.

5aSC16. Word timing in a counting task and perceived severity of impairment in spasmodicdysphonia. Masanobu Kumada (Dept. of Surgery, Yale Univ. School of Medicine, Haskins Labs., New Haven, CT 06511 and Dept. of Physiology, Natl. Defense Col., Japan), Fredericka Bell-Berti (St. John's Univ., Jamaica, NY and Haskins Labs., New Haven, CT), Takeo Kobayashi (Teikyo Univ., Ichihara Hospital, Tokyo) Kiyoshi Makiyama (Nihon Univ. Surugadai Hospital, Tokyo, Japan), and Seiji Niimi (Univ. of Tokyo, Tokyo, Japan)

Since the mora is the phonologically isochronic unit in Japanese, the proportion of impaired morae may be used as an indicator of the temporal impairment of speech. In spasmodicdysphonia (SD), symptoms occur sporadically and at three levels: (1) voice quality, (2) phonological, and (3) fluency. In previous work [Kumada *et al.*, Jpn. J. Logoped. Phoniatr. **38**, 176–181 (1997)], the ''Mora Method'' was proposed to evaluate the severity of SD in Japanese-speaking patients, with the severity represented by the proportion of impaired morae in a 25-morae sentence. Since counting aloud is a fairly rhythmic task, this paper studies using a proportion of impaired syllables in a counting task to characterize the severity of impairment in English-peaking SD patients. In addition to perceptual syllable-by-syllable judgments, the paper also examines timing relations of number-to-number and syllable-to-syllable intervals for comparison with syllable-impairment and overall-quality rating scale judgments. These comparisons should help establish the usefulness of perceptual syllable-by-syllable judgments to represent the severity of fluency impairment in English-speaking SD patients. [Work supported by NIH DC-00865 to Haskins Laboratories and by St. John's University.]

5aSC17. LF source model adequacy for pathological voices. Brian Gabelman, Jody Kreiman, Bruce R. Gerratt, Norma Antonanzas-Barroso (Div. of Head and Neck Surgery, UCLA School of Medicine, 31-24 Rehab. Ctr., Los Angeles, CA 90095-1794, jkreiman@ucla.edu), and Abeer Alwan (UCLA, Los Angeles, CA 90095)

The adequacy with which source models describe pathological phonation has been little studied, and the perceptual importance of deviations from a perfect model fit remains unknown. To address these issues, pathological voices were selected randomly from a library of samples. Glottal source waveforms were estimated by inverse filtering airflow signals. Formant frequencies and bandwidths were estimated using combined LPC and spectrographic analyses. Source functions were synthesized by least-squares fitting a simplified LF model [Y. Qi and N. Bi, J. Acoust. Soc. Am. **96**, 1182–1185 (1994)] to the glottal flow derivative. Three versions of each voice were created. The first was constructed by extracting one cycle from the original voice and repeating it to form a 1-s signal. The second was constructed by recombining the inverse filtered source and estimated vocal tract resonances. The third version comprised the LF-modeled source combined with the estimated vocal tract resonances. Expert listeners compared the three versions of each voice, and results were interpreted in terms of the specific differences among stimuli. This procedure allowed evaluation of the extent and perceptual importance of information lost first by inverse filtering and then by fitting the LF model to the resulting volume velocity signal. [Research supported by NIH.]

5aSC18. Constraints on syllable structure in speech. Jeffery A. Jones, Kimberly Purdy, and Kevin G. Munhall (Dept. of Psych., Queen's Univ., Kingston, ON K7L 3N6, Canada, jones@pavlov.psyc.queensu.ca)

This paper presents kinematic data and a statistical analysis of the MRC Psycholinguistic Database in a study of patterns of syllable coda structure in English bisyllables. X-ray microbeam and OPTOTRAK data were used to study spontaneous errors in rapid speech. Subjects repeated real-word and nonsense bisyllables at increasing speaking rates. All subjects made errors that harmonized the place of articulation of the codas at faster speaking rates. Errors were determined by kinematic changes in the tongue and lip gestures. Analysis of the MRC Psycholinguistic Database revealed that coda harmonies occur statistically more frequently than chance in English even when the statistical incidence of segments are taken into account. The experimental evidence is discussed in terms of production constraints on syllable production. The influence of stress pattern, manner, and place of articulation are explored. [Work supported by NIH-NIDCD Grant No. DC-00594 and NSERC.]

5aSC19. Kinematic and spectral measures of supralaryngeal correlates of the accent contrast in Australian English high vowels. Jonathan Harrington (Speech Hearing and Lang. Res. Ctr., Macquarie Univ., Sydney, Australia), Mary E. Beckman (Ohio State Univ., Columbus, OH), Janet Fletcher (Melbourne Univ., Melbourne, Australia), and Sallyanne Palethorpe (Macquarie Univ., Sydney, Australia)

Six talkers of Australian English read 108 dialogues that induced accented or unaccented renderings of target bisyllabic words in which the rhythmically strong initial syllable contained a tense or lax high vowel flanked by bilabial, alveolar, or velar consonants. The movement of the tongue dorsum and jaw were tracked in horizontal and vertical dimensions using an electromagnetometer system, and synchronized recordings were made of the acoustic signal. The duration of the target words was usually greater in accented position in the dialogue. However, for most of the talkers the effect was confined to the initial consonant. For many talkers, the accented vowel was acoustically more peripheral and was produced with a greater fronting of the tongue dorsum, with differences in fronting extending from the vowel's acoustic onset to its target. However, the magnitude of the effect of accent on supralaryngeal displacement was, in general, considerably less than the influence exerted by the consonantal context and there was no evidence that accent diminishes segmental coarticulatory influences. These results suggest that supralaryngeal cues to the accentuation of a vowel are strictly relational in the sense that they require a decoding of the segmental context in which they occur.

5aSC20. Rate manipulation and stability of onset and coda structures. Kenneth J. de Jong (Dept. of Linguist., 322 Memorial Hall, Indiana Univ., Bloomington, IN 47403, kdejong@indiana.edu)

Stetson (1951) noted that singleton coda consonants, such as in the word ''eat,'' will shift to onset consonants, such as in the word ''tea,'' when the syllable is repeated at fast rates. Tuller and Kelso (1991), replicating Stetson's effect, claim glottal phasing to be the relevant collective variable for examining syllable structures, and posit a dynamical model in which onsets represent a mode of coordination into which the articulatory system gets attracted. The present paper reports an experiment wherein syllables consisting of /i/ and labial stops are elicited in various rate-changing conditions. Speech rate is controlled with a metronome. Results partially support Tuller and Kelso's model in that coda /p/ and /b/ show shifts in voice onset time to the stable values (expressed in absolute milliseconds) exhibited by onsets under increases in speech rate. In addition, results show hysteresis on the rate at which these shifts occur. However, the data show that coda configurations are also stable. When occlusion duration is measured, codas are stable across rates, while onsets are instable, shifting from longer values into the stable values exhibited by the codas. Thus, both onset and coda structures represent stable configurations, though of very different types.

5aSC21. Some factors influencing duration between syllables judged perceptually isochronous. Laura Dilley (Speech Commun. Group, MIT, 50 Vassar St., Cambridge, MA 02139)

This pilot study examines the influence of several factors on time intervals between syllables judged to be isochronous, including number of intervening segments and number of intervening syllables. Earlier work [Fant *et al.* (1991); Lehiste (1977)] has primarily been concerned with interstress intervals, rather than perceptual ''beats,'' though more recent work has examined beats per se [Couper–Kuhlen (1993)]. For our study,

five labelers with musical backgrounds applied a system for labeling perceptual isochrony to 1 min each of spontaneous, read, and radio news style speech [Dilley, forthcoming]. For strings indicated as isochronous by three or more labelers we determined the intervals between the ''rhythmic beats,'' i.e., perceptually strong syllables which formed the perceived isochronous sequences. For simplicity, the ''beat length'' was taken to be the time between successive vowel midpoints (or syllabic sonorant segments) for these perceptually strong syllables. Preliminary results for 1.5 min of speech indicate that beat lengths range from 200 to 800 ms, and that there is considerable variability for successive beat lengths in a perceptually isochronous sequence. This variability can be reduced somewhat by adjusting for the number of intervening segments.

5aSC22. Difference limens for frequency and amplitude perturbation as a function of fundamental frequency and waveshape. Shari L. Campbell (Dept. of Commun. Sci. and Disord., Univ. of Georgia, 514 Aderhold Hall, Athens, GA 30602, scampbel@coe.uga.edu)

Three-interval forced-choice adaptive procedures were used to estimate difference limens (DLs) for signals with either frequency perturbation (jitter) or amplitude perturbation (shimmer). A time-domain synthesis technique was used to generate sinusoidal and triangular waves with specified amounts of both perturbation types (defined as the mean absolute value, in percent, of the first-order perturbation function) for each of two mean fundamental frequencies (100 and 200 Hz). Standard stimuli had neither frequency nor amplitude perturbation. Sequential characteristics of the signals were determined via autocorrelation analyses. Signals were 900 ms in duration and were presented monaurally at 80 dB SPL via insert earphones to ten listeners with normal hearing. Up to 12 estimates were obtained for each of the eight signal types. Group average DLs ranged from 0.07% (frequency perturbation, 200-Hz triangular wave) to 6.20% (amplitude perturbation, 100 Hz triangular wave). For a given fundamental frequency and waveshape, DLs were smaller for frequency perturbation than for amplitude perturbation. Within a perturbation type, DLs were smaller for triangular waves with frequency perturbation and for sine waves with amplitude perturbation. [Work supported by The University of Georgia Research Foundation.]

5aSC23. Insights on normal vocal-fold oscillation from the acoustic dissection of single glottic pulses in vocal fry. S. A. Elder (Phys. Dept., 572 Holloway Rd., U.S. Naval Acad., Annapolis, MD 21402) and P. F. Castellanos (Univ. of Maryland School of Medicine, Baltimore, MD 21201)

Vocal-fold oscillation has been investigated from the viewpoint of kinematics and dynamics of single glottal pulse at normal intensity vocal fry frequencies, using a modified Sondhi tube. Key moments in the cycle are opening and closing actions, which appear to be very sudden. These events may be noted in two ''kinks'' which appear along traces of the EGG and glottal cross-sectional area functions. Between opening and closing boundary encounters, vocal folds seem to travel in quasiballistic fashion, being slingshotted out by feedback from the subglottal pressure wave and ending their traverse across the glottal opening with mutual collision The abrupt opening and closing gestures of the glottis permit the glottal valve to modulate subglottal volume flow against the inertia of tracheal air. Acoustic emissions from the glottis may be thought of as flow fluctuations which compensate for changes in fluctuating glottal area, thereby maintaining approximately incompressible flow. Fluctuations of glottal area generate the bulk of F0 energy while upper formants owe their spectral energy to boundary layer events that happen near the violent endpoints of glottal motion.

5aSC24. Acoustical evidence for the role of preaspiration in Norwegian. Bente H. Moxness (Dept. of Linguist., Norwegian Univ. of Sci. and Technol., 7055 Trondheim, Norway, benmox@alfa.itea.ntnu.no)

The Norwegian quantity system exhibits two main rhyme types; V[long]C[short] and V[short]C[long]. In addition, the two rhyme types may contain both voiced and voiceless consonants. Preaspiration is a scarce phenomenon in the world's languages, but it is found in varying degrees in Norwegian dialects. A pilot study showed no preaspiration in rhymes containing voiced consonants. Acoustical evidence is presented as documentation for the presence of preaspiration in the production of Norwegian rhymes containing voiceless consonants. Through looking at preaspiration as a relative value of overall rhyme duration, it is shown that preaspiration is significantly greater in V[short]C[long,voiceless] rhymes compared to V[long]C[short,voiceless] rhymes. Thus preaspiration may be considered a cue to V[short]C[long,voiceless] rhymes in Norwegian. A gender difference was noted as female speakers exhibited more preaspiration than male speakers in V[short]C[long,voiceless] rhymes, perhaps due to breathier voice qualities among females. It is assumed that preaspiration in Norwegian has diminished through the development from Old Norse into modern Norwegian, and a perception study shows that preaspiration holds no perceptual value in today's spoken Norwegian. The Norwegian quantity system is compared to the Icelandic quantity system, which resembles Old Norse. The Norwegian dialect used in the study is the Trønder dialect.

5aSC25. Voice *F*0 responses to feedback pitch are not affected by speed of stimulus onset. Theresa A. Burnett, Swathi Kiran, and Charles R. Larson (Dept. of Commun. Sci. and Disord., Northwestern Univ., 2299 N. Campus Dr., Evanston, IL 60208)

Voice *F*0 appears flexibly controlled by interactions between voluntary mechanisms, memory, and auditory feedback. Previous studies have shown that voice *F*0 is changed following an abrupt shift in feedback pitch [Elman, 1981; Kawahara, 1994; Larson *et al.*, 1995]. In order to establish that this response is due to the pitch of the feedback shift rather than its onset speed, an experiment was conducted in which voice feedback pitch heard loudly through headphones was slowly or quickly decreased to 100 cents lower than the steady voice *F*0 produced by the subjects. Subjects were instructed to maintain steady *F*0 production regardless of any feedback changes they might hear. Both slow and fast ramping pitch shift stimuli (PSS) generally elicited opposing responses characterized by a rising voice *F*0 which mirrored the slope of the lowering PSS ramp. Accordingly, response latency and time of peak response were inversely related to the slope of the PSS ramp; the lower the slope, the later the response latency and peak. Peak response magnitude was not affected by ramp speed. These results suggest that the auditory–vocal system continuously monitors auditory feedback and adjusts vocal output so as to reduce the difference between it and a reference signal.

5aSC26. Characteristics of roughness in esophageal speech. Hector Javkin, Nancy Niedzielski, and James Reed (Panasonic Technologies, Inc., Speech Technol. Lab., 3888 State St., Santa Barbara, CA 93105, hj@stl.research.panasonic.com)

Esophageal voice, produced by the vibration of the superior sphincter of the esophagus by persons who have been laryngectomized, is almost invariably associated with voice source roughness. Using inverse filtering of speech recorded to minimize phase distortion, the source characteristics of six esophageal speakers with different levels of proficiency were examined. With our three less proficient speakers, the roughness is clearly evident in the volume velocity waveform. Highly proficient esophageal speakers, however, although they have a voice quality which is perceived as rough, nevertheless produce a volume velocity waveform surprisingly similar to that of laryngeal speakers. The characteristics that produce this perceived roughness, other than jitter and shimmer, are difficult to detect in the volume velocity waveform. They are noise components of low

amplitude and relatively high frequency, which are attenuated in the (integrated) volume velocity. They are best observed in the differentiated volume velocity waveform or in spectral measures. Our results suggest that the volume velocity waveform may not be a very effective measure for evaluating the speech of highly proficient esophageal speakers.

5aSC27. Modeling and analysis of speakers with vocal-fold nodules. Jeff Kuo (MIT Res. Lab. of Electron., Rm. 36-511, Cambridge, MA 02139, kuo@speech.mit.edu), Robert E. Hillman (MGH—Inst. of Health Professions, Boston, MA), and Eva B. Holmberg[a] (Massachusetts Eye and Ear Infirmary, Boston, MA)

Speakers with vocal-fold nodules commonly use higher than normal subglottal pressures to produce a particular sound pressure level (SPL). The SPL may be low because of a decreased maximum flow declination rate (MFDR) of the glottal flow waveform, increased first formant bandwidth, or increased spectral tilt. A modified two-mass model of vocal-fold vibration is proposed, which suggests that the MFDR can be decreased because of increased coupling stiffness between the masses, altered equilibrium positions for the masses, and the presence of nodules that interfere with the normal closure activity of the vocal folds. A parallel acoustic and aerodynamic study of selected patients with vocal nodules was conducted. Preliminary findings show higher than normal calibrated amplitudes of the first harmonic ($H1$) normalized by the SPL, as well as larger amplitude differences between the first harmonic and first formant ($H1-A1$). These results agree with the previous finding of a larger ac component of the glottal flow waveform [Hillman *et al.*, J. Speech Hear. Res. **32**, 373–392 (1989)]. The presence of a glottal chink may also widen the first formant bandwidth. Relationships between modeling results and experimental data will be discussed. [Work supported by NIH Grants DC00075 and DC00266.] [a] Presently at Stockholm Univ., Sweden.

5aSC28. Neonates attend to perceptual cues to rudimentary grammatical categories. Rushen Shi and Janet Werker (Psych. Dept., Univ. of British Columbia, Vancouver, BC V6T 1Z4, Canada)

To acquire the rules by which words are combined to form phrases and sentences, infants must first learn to assign words to grammatical categories such as nouns, prepositions, etc. This study is based on the assumption that infants initially derive the binary, superordinate lexical and functional categories from perceptual cues in the input. These two categories are present in all human languages, and recent work reveals similar distinctive acoustic and phonological cues across languages [R. Shi, J. Morgan, and P. Allopenna, J. Child Language (in press)]. For this study it was therefore proposed that there exists an innate mechanism leading to initial sensitivity to cues separating the two universal categories. Newborns were tested using the high-amplitude sucking paradigm. Stimuli were multiple lexical and function word tokens randomly selected and sliced from natural infant-directed utterances of an English-speaking mother. After being habituated to a word set of one category, infants were tested on either words of the other category or new words from the same category. The results showed that infants' recovery was significantly greater to words of the other category. This evidence supports a predisposed sensitivity which may serve as an initial foundation for the formation of early grammatical categories.

5aSC29. Interference in short-term pitch and timbre memory due to the structure of a tone sequence. Gary E. Starr (Wingate Univ., Campus Box 5005, Wingate, NC 28174, gstarr@wingate.edu)

A number of studies have investigated memory for pitch and timbre using a tone discimination task [e.g., G. Starr and M. Pitt, J. Acoust. Soc. Am. **102**, 486–494 (1997)], where standard and comparison tones are separated by a retention interval filled with interference tones. These studies found same-dimension similarity to be the primary source of interference. However, the organizational structure of the tone sequence may also affect interference. The present project used two different tasks to evaluate the role of organization on interference in memory for pitch and timbre. The first task was identical to that of previous research. In the second task, listeners discriminated either the pitch or timbre of standard and comparison tones when the standard was presented as the sixth tone of an 11-tone sequence. In both tasks, sequences were constructed that varied the organizational structure of the sequences, while same-dimension similarity was kept constant. The organizational structure of the tone sequences was found to influence interference in both tasks, suggesting that factors other than same-dimension similarity disrupt memory.

FRIDAY MORNING, 5 DECEMBER 1997 FRIARS/PADRE ROOM, 8:30 TO 11:00 A.M.

Session 5aSP

Signal Processing in Acoustics: Microphone Arrays

Douglas Meegan, Chair
Applied Research Laboratories, University of Texas, P.O. Box 8029, Austin, Texas 78713-8029

Contributed Papers

8:30

5aSP1. Microphone array sensor placement optimization in reverberant environments. Daniel V. Rabinkin, Richard J. Renomeron, and James L. Flanagan (Ctr. for Comput. Aids for Ind. Product., Rutgers Univ., P.O. Box 1390, Piscataway, NJ 08855-1390, rabinkin@caip.rutgers.edu)

Microphone arrays can be used for high-quality sound pickup in reverberant and noisy environments. The beamforming capabilities of microphone array systems allow highly directional sound capture, providing a superior signal-to-noise ratio (SNR) when compared to single microphone performance. Recent work [Rabinkin *et al.*, ''Optimum microphone placement for array sound capture,'' Proc. SPIE 8/97] has addressed the issue of microphone placement for optimized array performance. A Monte Carlo procedure was described to evaluate average sound capture SNR for a given microphone configuration based on geometry-related array performance statistics. Numerical optimization was performed for particular classes of sensor geometries based on the evaluated SNR. A uniformly distributed uncorrelated additive noise model was used to evaluate performance. An improved performance model is described which accounts for the correlated nature of additive noise generated by acoustic reflections of

5a FRI. AM

the source in a reverberant environment. The Allen and Berkley image method is used to generate the reflective noise portion. Optimization is performed for various acoustic conditions and array performance is compared for different reverberation levels.

8:45

5aSP2. Small-scale matched filter array processing for spatially selective sound capture. Richard J. Renomeron, Daniel V. Rabinkin, Joseph C. French, and James L. Flanagan (Ctr. for Comput. Aids for Ind. Product., Rutgers Univ., P.O. Box 1390, Piscataway, NJ 08855-1390)

Matched filter array (MFA) processing has a distinct advantage over delay-sum beamforming for reverberant enclosures in that it is able to cohere significant reflected images in addition to the direct arrivals and is capable of spatial selectivity in three dimensions. However, previous study of this technique utilized a very large number of sensors, which is beyond the A/D capacity and computational power of current commercially available DSP hardware. In this paper, the results of using matched filtering with only eight sensors in real rooms are presented. It is found that a moderate amount of reverberation can be removed, which suggests that the technique may prove useful for high-quality speech pickup for teleconferencing in a small-room environment.

9:00

5aSP3. A beamforming paradox: Using far-field techniques to design a near-field microphone array. Darren B. Ward (Acoust. Speech Res. Dept., Bell Labs, Lucent Technologies, Murray Hill, NJ 07974, dbward@research.bell-labs.com), Rodney A. Kennedy, and Thushara Abhayapala (Australian Natl. Univ., Canberra, ACT 0200, Australia)

A new method of designing a beamformer having a desired near-field beampattern is presented. Although a plethora of techniques have been proposed for far-field beamformer design (where the received wavefronts are planar), these methods generally result in severe beampattern degradation when used in a near-field environment (when the source is sufficiently close to the array that the spherical nature of the wavefronts must be taken into account). In this paper a computationally simple design procedure is presented in which established *far-field* design techniques can be employed to achieve *near-field* beampatterns with high accuracy. The theoretical underpinnings of this procedure are based on the spherical solution to the wave equation. The method can be applied to the problem of designing a microphone array for speech acquisition from a near-field talker.

9:15

5aSP4. An efficient method achieving high-precision digital time-shifting. Chen Liu (3065 Beckman Inst., Univ. of Illinois, Urbana, IL 61801, liu2@uiuc.edu)

Time shifting is an important procedure required in acoustic applications and research such as producing interaural time difference in auralization for audio virtual reality and simulating multiple inputs for spatial sound signals in research on microphone arrays. Traditional analog time-shifting methods necessitated electronic delay lines, which were a burden on hardware. Instead, more recent digital methods performed time shifting using software. However, the digital methods required an extraordinarily high-sampling rate in order to achieve a reasonable precision in time shifting. Most of the digital methods so far employed interpolation to cope with this problem, although interpolation incurred a great amount of extra computation. A high-precision frequency-domain time-shifting method was recently developed for the purpose of digital implementation of broadband beamformers [Liu and Sideman, J. Acoust. Soc. Am. **98**, 241–247 (1995)]. In this study, the above method was extended to more general application contexts. The interpolation was employed implicitly in this method. The computation was required only for doing FFT and unrelated with interpolation. Theoretically, the time-shifting precision can be arbitrarily high whereas the amount of computation was kept constant. Moreover, this method can easily be implemented in real time by means of the conventional short-term Fourier transform.

9:30

5aSP5. Enhanced target identification using co-located acoustic and seismic sensors. Mark C. Wellman (U.S. Army Res. Lab., Adelphi, MD 20783), Nassy Srour, and Douglas E. Lake (U.S. Army Res. Lab., Adelphi, MD 20783)

Feature extraction and fusion of target signatures is of great importance in the detection, tracking, and identification of tracked and wheeled vehicles. Current research is centered in developing a set of robust features which will allow classification with high confidence regardless of various environmental conditions. The performance of boundary decision and nonparametric classifier topologies is also of special interest for the acoustic/seismic feature space. Acoustic and seismic data collected using co-located sensors have been processed using harmonic line association [J. A. Robertson, IIT Research Institute, in-house report] and higher order statistic [F. B. Shin and D. H. Kil, Proceedings of the International Conference on Image Processing, **3** (1995)] feature extraction techniques for classification. Performance results of the feature fusion will be presented as well as several classifier architectures. Real-time performance will also be presented [N. Srour, Remote Netted Acoustic Detection System, Army Res. Lab., May (1995)].

9:45–10:00 Break

10:00

5aSP6. A broadband model-based array processor. Edmund J. Sullivan (Code 82101, Naval Undersea Warfare Ctr., 1176 Howell St., Newport, RI 02841) and James V. Candy (Lawrence Livermore Natl. Lab., Livermore, CA 94551)

Model-based processing for towed arrays is based on a Kalman-type estimation scheme. In the narrow-band case the measurement equation is based on either a plane wave or curved wavefront monochromatic signal model. In the broadband case one could model the signal as a set of parallel narrow-band signals. However, this would be prohibitively computationally intensive. Here, it is shown how the signal can be modeled with very few parameters using an autoregressive moving average (ARMA) model for the signal, i.e., the measurement equation in the Kalman filter, where the parameters of interest, the bearing and range, are combined with the ARMA parameters in a single state vector, thus providing an adaptive scheme wherein these parameters are adapatively estimated in an iterative manner. Examples of broadband bearing and range estimation based on simulated data will be shown.

10:15

5aSP7. Semicoherent technique to improve sonar processor performance for signals with small fluctuations. Ronald A. Wagstaff and Jackson Mobbs (Naval Res. Lab., Stennis Space Center, MS 39529-5004, wagstaff@nrlssc.navy.mil)

A signal processing algorithm that utilizes coherent addition of the acoustic pressures usually provides greater noise suppression with a corresponding increase in signal-to-noise ratio (SNR) than one that uses incoherent addition. Furthermore, fluctuation based processing (FBP) algorithms that are designed to operate on signals from submerged sources (i.e., small fluctuations) generally utilize incoherent averaging techniques to provide various enhancements over conventional algorithms that are also incoherent, including greater SNR, increased resolution, clutter suppression/elimination, and unalerted auto-detection/classification (submerged/not-submerged). An algorithm that combines the desirable features of the coherent approach and the incoherent FBP approach could possess desirable features of both types of algorithms and be very useful for enhancing the processor's performance. Such an algorithm has been devised. It operates on the combined amplitude and phase fluctuations in the acoustic pressures in a semicoherent manner to substantially improve the processor's performance. The algorithm will be defined, and results from measured undersea data will be presented to illustrate the gains that are achieved by this new algorithm. [Work supported by ONR and NRL.]

10:30

5aSP8. AWSUM III_{434}: A coherent fluctuation based processor for enhanced resolution and unalerted autodetection. Ronald A. Wagstaff and Jacob George (Naval Res. Lab., Stennis Space Center, MS 39529-5004, wagstaff@nrlssc.navy.mil)

A new coherent fluctuation based processor for enhancing spatial resolution, suppression of clutter, and unalerted autodetection, designated AWSUM III_{434}, is described and results are presented. The processor is a combination of the Wagstaff's Integration Silencing Processor (WISPR) III [R. A. Wagstaff and J. George, SPIE Aerospace/Defense Sensing and Controls '96 Symposium, Targets and Background: Characterizations and Representation II Conference, Proc. SPIE **2751**, 132–141 (1996)] and the Advanced WISPR Summation (AWSUM) processor [R. A. Wagstaff, IEEE J. Ocean. Eng. **22**, 110–118 (1997)]. While WISPR III achieves signal-to-noise ratio (SNR) gains of about 14 dB, AWSUM III_{434} achieves gains of about 24 dB. In addition, the spatial resolution of AWSUM III_{434} is significantly enhanced, and the unaltered autodetection/identification capability for signals from submerged sources is measurably increased, e.g., by 10 dB. [Work supported by ONR and NRL.]

10:45

5aSP9. Unalerted autodetection and other signal processor enhancements by exploiting phase fluctuations. Ronald A. Wagstaff and Jackson Mobbs (Naval Res. Lab., Stennis Space Center, MS 39529-5004, wagstaff@nrlssc.navy.mil)

Fluctuations in signals and noise are an inherent characteristic of the undersea acoustic environment. The effects of the fluctuations were usually considered a nuisance to be avoided or ignored. However, within the last few years, it has been recognized that the fluctuations in the acoustic power amplitudes received by a sonar system could be exploited to achieve additional signal-to-noise ratio (SNR) gains, enhanced resolution, clutter suppression and elimination, and unalerted autodetection/classification. This paper demonstrates that a similar capability can be achieved by exploiting the phase fluctuations in the acoustic pressures of the signals and the noise. In fact, since the phase is independent of the pressure amplitude, the phase-dependent algorithm is able to enhance the sonar's performance on some phase-stable signals, which have fluctuation amplitudes that will cause the other techniques that exploit the amplitude fluctuations to fail. The new technique for exploiting phase fluctuations is described, and results from measured data are presented that show SNR gains of more than 6 dB, complete clutter elimination, and unalerted automatic detection/classification (submerged/not submerged) of signals from submerged sources. [Work sponsored by ONR and NRL.]

FRIDAY MORNING, 5 DECEMBER 1997 CALIFORNIA ROOM, 8:00 TO 11:45 A.M.

Session 5aUW

Underwater Acoustics: Bottom Interaction and Scattering

Gregory J. Orris, Chair
Naval Research Laboratory, 4555 Overlook Avenue, Southwest, Washington, DC 20375-5000

Contributed Papers

8:00

5aUW1. Dispersion of sound in a sandy silt sediment. Nicholas P. Chotiros (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

Sound pulses penetrating from water into a sandy silt were found to be dispersive in the 1- to 10-kHz band. Pulses were projected from a parametric source into a sandy silt sediment and received by a buried array of hydrophones. The signals were coherently processed for direction and speed of the sound pulse as a function of frequency. The sound speed was found to increase with frequency at a rate greater than that expected from simple attenuation. Acoustic propagation was modeled with Biot's theory using constant coefficients, and model predictions were unable to match the observed dispersion. One possible explanation is a frequency-dependent frame bulk modulus, due to relaxation effects associated with the interaction between the pore fluid and the grain-grain contacts. [Work supported by ONR 321 OA.]

8:15

5aUW2. Measuring the locations of penetrating acoustic ray paths into marine sediments. Frank A. Boyle and Nicholas P. Chotiros (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

Recent experiments at sea and in the laboratory indicate anomalously high levels of acoustic penetration into sandy marine sediments at subcritical grazing angles, which are not accounted for via current viscoelastic acoustic models for sediments. Current research is directed toward a possible physical explanation. One possibility [Chotiros, J. Acoust. Soc. Am. **97**(1), 199–214 (1995)] is that the sediment acts as a Biot poroelastic medium, supporting Biot slow waves with sufficient amplitude to account for the observed penetration. Another [Thorsos *et al.*, SACLANT Conf. Proc. CP-45, 563–569 (1997)] is that seabed interface roughness scatters a significant amount of sound into the sediment. These two mechanisms differ in their ray geometries for penetrating sound. In order to determine the relative importance of the two mechanisms, a tank experiment was performed to determine acoustic ray paths into a sandy bottom. It involved a narrow beam acoustic source directed at hydrophones in the sediment column. Initial results appear to be consistent with Biot slow-wave penetration. A description of the experimental work will be included, along with a discussion of preliminary results.

8:30

5aUW3. Evaluation of Biot–Stoll theory for conversion of geology to geoacoustic data in shallow water. Mohsen Badiey (College of Marine Studies, Univ. of Delaware, Newark, DE 19716), Alexander H. D. Cheng, and Yongke Mu (Univ. of Delaware, Newark, DE 19716)

A procedure for estimating acoustic-wave velocity and attenuation in ocean sediment using a minimum amount of geological and geotechnical data is demonstrated. First, the Biot–Stoll theory is presented. Next, various asymptotic formulas for the attenuation coefficient are derived for high, low, and intermediate frequencies. These expressions clearly isolate the effects of the intergranular Coulomb friction and fluid viscous dissipation on the attenuation of shear and compressional waves. Under the constraint of a minimum amount of geological and geotechnical information, a sequence of empirical equations is compiled to convert basic data, such as the blow count number from a standard penetration test or ship-

board density, into sediment geoacoustic properties. As a demonstration, two well-known field cases, the Atlantic Generating Station (AGS) site and the Atlantic Margin Coring (AMCOR 6010) site, are examined. By incorporating the uncertainty involved in the data collection, the estimated geoacoustical parameters are provided with a standard deviation.

8:45

5aUW4. A synthetic array measurement of a fast compressional and a slower wave in an unconsolidated water-saturated porous medium. Harry J. Simpson and Brian H. Houston (Naval Res. Lab., Code 7136, 4555 Overlook Ave., S.W., Washington, DC 20375-5350, simpson@fugazi.nrl.navy.mil)

An investigation of sound waves propagating in an unconsolidated water-saturated porous medium has been conducted in the NRL shallow-water laboratory. To minimize the interaction of the water–sand interface, a transducer was buried approximately 1.5 m facing up toward the sand–water interface. The unconsolidated water-saturated porous medium is a 3.0-m deep bed of a manufactured 212-μm mean diameter sand. To obtain a short pulse that provides better temporal resolution of the acoustical waves propagating in the media, a 1-3 composite transducer was used. The sound field is measured with a single hydrophone that is initially buried just above the source and then incrementally pulled up through the sand toward the sand–water interface. This provide a synthetic array meaurement of the sound field directly above the transducer. A 1680-m/s fast compressional wave and a 550-m/s slower wave are easily observed using both temporal and spatial analysis techniques. Comparisons of these wave speeds with predictions from a Biot model are discussed. [Work supported by ONR.]

9:00

5aUW5. The enormous acoustic nonlinearity of air-contained sand sediments. Dimitri M. Donskoy (Davidson Lab., Stevens Inst. of Technol., Hoboken, NJ 07030, ddonskoy@stevens-tech.edu) and Alexander Sutin (Stevens Inst. of Technol., Hoboken, NJ 07030)

The nonlinear acoustic properties of sediments have been studied using a novel experimental approach. A small amount (10–100 cm^3) of sediments is suspended in a water column in a sound transparent vessel. The vessel is placed inside an experimental tank and exposed to harmonic acoustic pressure. Nonlinearity of sediments causes the generation of higher harmonics which are radiated into the surrounding water and measured with a hydrophone. The nonlinear response for various frequencies (0.1–10 kHz) of the excitation is studied varying both the grain-size distribution of the sediment and the amount of trapped air. A very significant level of nonlinear distortion is observed even for a relatively low amplitude of the excitation signal (under 200 Pa). In some cases (coarse sand with a 0.01–0.03 gas void friction) up to 20 harmonics are observed, with the second harmonic attaining 30% (-10 dB) of the fundamental. Such a high level is 120 dB greater than the estimated amplitude of the second harmonic generated by the same volume of pure water and more than 40 dB greater than could be generated by water having the same gas void friction. The theoretical model of such a considerable nonlinear behavior based on the interaction of grains in the presence of gas voids is discussed. [Work supported by ONR.]

9:15

5aUW6. A class of low-frequency modes in range-independent fluid–solid media. Sven Ivansson (Div. of Underwater Wave Propagation and Active Sonar, Natl. Defence Res. Establishment (FOA), 17290 Stockholm, Sweden)

A unified treatment is given of the class of low-frequency P–SV modes whose complex phase velocities do not approach zero as fast as the frequency when the frequency is decreased toward zero. Utilizing the analyticity of the dispersion function, a complete characterization is given of these modes and the linearly viscoelastic fluid–solid media in which they occur. The two Lamb plate modes and the classical interface waves (Rayleigh, Scholte, Stoneley) all appear in a general setting, and asymptotic low-frequency expressions are given for modal slownesses and mode forms. The corresponding phase velocities will either tend to a nonzero constant or approach zero like the square root of frequency. In addition, slow modes appear whose phase velocities approach zero like three other powers of frequency: 1/3, 3/5, and 2/3. Concerning 1/3, a correction is made to a paper by Ferrazzini and Aki; 3/5 and 2/3 appear for certain bending-type waves that are slower than the classical bending wave. A less known type of interface wave also emerges from the analysis, and precise conditions for existence and uniqueness are obtained. The results are extended to interface conditions with slip. In an extreme case, yet another kind of interface wave is obtained.

9:30

5aUW7. Reflection coefficient of marine sediment covered with liquid pollutants. Henning Harms, Rainer Matuschek, and Volker Mellert (Dept. of Phys., Univ. of Oldenburg, D-26111 Oldenburg, Germany, rm@aku.physik.uni-oldenburg.de)

Acoustic sounding of sediment properties is usually carried out with waves at more or less normal incidence. This method, however, fails to detect thin layers ($d \ll \lambda$) of undissolved fluid covering the sea-bottom due to the small difference in impedance compared to water. But the spherical wave reflection coefficient is strongly affected at grazing incidence and allows for the detection of these layers. Laboratory measurements in model scale as well as model calculations are carried out in order to determine the capabilities of the new measuring method. Experiments were done in a water tank with a hydrophone as a spherical broadband sound source. Various liquid chemicals of density greater than water are poured on the sand in small amounts, forming layers of less than 3-mm thickness. Measured excess attenuation of the pure sediment is in very good agreement with simulations carried out with a fast field program (FFLAGS) that incorporates Biot theory [S. Tooms *et al.*, J. Acoust. Soc. Am. **93**, 173–181 (1993)]. The excess attenuation of the sediment coated by a thin layer differs significantly. In particular, the interference pattern is broadened and shifted toward lower frequencies. This behavior is explainable by FFLAGS as well. [Work supported by BMBF.]

9:45

5aUW8. Shallow-water bottom measurement with an L-shaped array. Yung P. Lee (Sci. Applications Int. Corp., 1710 Goodridge Dr., MS T1-3-5, McLean, VA 22102)

Bottom loss was measured at a shallow-water (188 m) site off the Gulf of Mexico using an L-shaped array. The L-shaped array consisted of a vertical component 153 m long covering most of the water column and a horizontal component 240 m long laid on the bottom. A source transmitting 10-s LFMs (40–440 Hz) continuously was towed toward, then away from the array. It was towed perpendicularly to the horizontal component of the array with a CPA (closest-point approach) about 250 m away from the center of the array. With the vertical component of the array, direct, surface reflected, bottom reflected, surface–bottom, bottom–surface, and surface–bottom–surface paths were identified. Bottom loss from single bottom bounce arrival over grazing angle from 10–40 deg were measured directly using the vertical component of the array. Because the horizontal component of the array was laid on the bottom, the ''(1+R) effect,'' the coherent sum of the arrival incident on the bottom and the resultant bottom reflection for that arrival, was measured. The (1+R) measurement was used to examine the geoacoustic bottom model of the site. [Work supported by the U.S. Navy.]

10:00–10:15 Break

10:15

5aUW9. The low-frequency radiation and scattering of sound from bubbly mixtures near the sea surface. William M. Carey (Dept. of Ocean Eng., MIT, 77 Massachusetts Ave., Cambridge, MA 02139) and Ronald A. Roy (Boston Univ., Boston, MA 02215)

Microbubble plumes are produced when waves break and are convected to depth. What role these microbubble plumes have in the production of sound and the scattering of sound near the sea surface from the low- (20-Hz) to mid- (2-kHz) frequency range? Ocean ambient noise

shows a dramatic increase in midfrequency levels when wave breaking occurs. Measurements of scattering from the sea surface have a different characteristic than expected by Bragg scattering from gravity waves, i.e., a large zero Doppler component. If microbubble clouds and plumes with void fractions greater than 0.0001 act as collective resonant oscillators, then noise can be produced and scattering can occur with little Doppler shift but ample spread. This hypothesis was based on the theory that the mixture properties determine the radiation and scattering from such a compact region. Experimental results show that the far field radiation of sound from a compact region can be described by monopole volume pulsation beneath a pressure release surface with the natural frequency described by a modified Minneart formula. Scattering measurements from submerged bubble clouds show a significant low frequency scattering. Experimental evidence and theoretical formulations are consistent with collective phenomenon. [Work supported by ONR.]

10:30

5aUW10. High-frequency bubble scattering theory and array processing. William M. Carey (Dept. of Ocean Eng., MIT, 77 Massachusetts Ave., Cambridge, MA 02139) and Ronald A. Roy (Boston Univ., Boston, MA 02215)

Microbubbles near the surface of the sea and in the water column can be an important factor limiting acoustic images in shallow water and near the surf zone. Classical theories exist which describe the scattering of sound from individual bubbles and sparse distributions of bubbles provided the damping factors are known. Backscatter and forward-scatter coherence loss can be described for random bubble distributions which are both uniform and stationary in space and time, and this results in phase correlation functions, distances, and variances. The spatial bubble population can be described by the statistical distribution of the turbulent field. This paper briefly outlines the relevant theory and applies it to free field bubble distributions at high frequencies. The impact on array processing is considered for a representative bubble size and spatial distributions. The expectation of the mean-squared array response is shown to depend on a coherence function and a coherence length proportional to the ratio of the phase correlation distance and variance which is related to the bubble spatial distribution. Extrapolations to what could be expected to be a limiting array resolution are presented for high-frequency side looking and synthetic aperture sonars. [Work supported by ONR.]

10:45

5aUW11. The interaction of acoustic signals with objects imbedded in sediment. Michael F. Werby (Code 7181, NRL, Stennis Space Center, MS 39529) and Natalia A. Sidorovskaia (Univ. of New Orleans, New Orleans, LA 70148)

The formulation of a method to describe the influence that an imbedded object has on an acoustic signal in a waveguide has taken on considerable interest not only for detection of man-made objects, but also in the study of inclusions in the ocean bottom that may have some geophysical significance. In this work, a method that allows one to make theoretical descriptions of quite general form for this kind of problem is considered. Some presented applications illustrate, where the influence of imbedded objects on a received acoustic signal is expected to be observed, and some conditions for which any influence is not observed. [Work sponsored by ONR, the Naval Research Laboratory, and the University of New Orleans.]

11:00

5aUW12. A high-order boundary-integral method for acoustic scattering by 3-D bodies in a layered fluid–solid medium. Ilkka Karasalo (FOA, Stockholm, Sweden) and Johan Mattsson (MWL, KTH, Stockholm, Sweden)

A boundary-integral equation (BIE) method is described for computing the acoustic field scattered by a 3-D object with a smooth, optionally penetrable, boundary in a layered fluid–solid medium. The medium is assumed to be horizontally homogeneous, with the extreme layers consisting of homogeneous rigid or penetrable half-spaces. The incident field is excited by a monofrequent point or line source, and the Green's function of the layered medium is computed by an adaptive transform integral technique. The BIE is solved by a point-collocation technique using high-order B splines (BSP) and high-order numerical integration. The discretized equations are solved by an iterative method, enhanced by preconditioning with the LU factors of a related linear system for spline interpolation or spline smoothing on the scatterer surface. Numerical examples are given at frequencies of a few kHz, representative, e.g., for a seabed-penetrating parametric transducer. A benchmark example is included, in which the field from a scatterer with super-ellipsoidal shape in a homogeneous medium is computed with the BSP method, with a standard BEM technique with quadratic elements, and with the null-field (NFL) method. The BSP method is shown to be highly efficient with performance gains growing as frequency increases and/or accuracy requirements are tightened.

11:15

5aUW13. Acoustic scattering by a three-dimensional elastic obstacle near a rough fluid–fluid interface. Raymond Lim (Coastal Systems Station, Code R22, 6703 West Hwy. 98, Panama City, FL 32407-7001), Kevin L. Williams, and Eric I. Thorsos (Univ. of Washington, Seattle, WA 98105)

The ensemble-averaged coherent and incoherent intensities are formulated for acoustic scattering by a smooth, bounded, elastic obstacle under the interface between two immiscible fluids with small-scale random roughness imposed. The formulation consists of combining Rayleigh-Rice perturbation theory for modeling propagation through the rough surface with a transition matrix solution for scattering by the obstacle near a planar surface. Results are applied to a spherical steel shell buried in a rough sediment bottom modeled with fluid parameters. It is demonstrated that the average incoherent field backscattered by the shell illuminated with shallow-grazing-angle acoustic sources can be well enhanced over field predictions based on scattering models where environmental surfaces are planar. Interestingly, the coherent intensity is also enhanced over flat interface predictions although not to the degree of the incoherent intensity. Implications to the acoustic detection and classification of buried obstacles are discussed. [Work supported by ONR.]

11:30

5aUW14. Scattering from fluid loaded targets in proximity to a sediment interface: Numerical results. Garner C. Bishop and Judy Smith (Naval Undersea Warfare Ctr. Div., Newport, RI 02840)

A null-field *T*-matrix formalism developed for plane-wave scattering from a fluid-loaded elastic spherical shell in proximity to a rough fluid elastic interface [J. Acoust. Soc. Am. **101**, 767–788 (1997)] is extended to include scattering from sound-hard and -soft spherical targets and a sound-hard finite cylinder with hemispherical end caps and, for a planar interface, sound-hard, sound-soft, and fluid-saturated poroelastic sediments. A variety of numerical results are obtained that demonstrate some of the effects of target and sediment parameters on target–boundary acoustic coupling and target strength. Then, for small amplitude surface roughness, a perturbation formalism is used to indicate some of the effects of interface roughness.

Session 5pAB

Animal Bioacoustics: Instrumentation in Animal Bioacoustics II

Mardi C. Hastings, Cochair
Department of Mechanical Engineering, Ohio State University, 206 West 18th Avenue, Columbus, Ohio 43210-1107

Charles R. Greene, Cochair
Greeneridge Sciences, Inc., 4512 Via Huerto, Santa Barbara, California 93110

Contributed Papers

1:00

5pAB1. Three-dimensional localizations of vocalizing blue whales. Gerald L. D'Spain, Aaron M. Thode, and William A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0704)

During a recent experiment in the Southern California Bight region, recordings of vocalizing blue whales were made simultaneously by a vertical line array of 48 hydrophones and freely drifting, standard Navy sonobuoys. Vector acoustic intensity techniques are applied to the ''DIFAR'' sonobuoy data to determine azimuth versus time for two individual animals vocalizing in turn. Once the azimuths are determined, matched-field processing with a 2-D normal-mode code is applied to the vertical hydrophone array data to determine the animals' position in range and depth. These processing techniques used with the data from these two sensor systems permit investigations to be made of diving behavior, vocalization source levels, and calling interactions. [Work supported by ONR, Codes 341BST and 321OA.]

1:15

5pAB2. Passive acoustic tracking of echolocating harbor porpoise (*Phocoena phocoena*). I. Fisheries management requirements and biological basis. David C. Potter, Janeen M. Quintal (Natl. Marine Fisheries Service, Woods Hole, MA 02543), James H. Miller, Robert Gampert (Univ. of Rhode Island, Narragansett, RI 02882), and Khosrow Lashkari (MBARI, Moss Landing, CA 95039)

Harbor porpoises are incidentally taken in various sink gillnet fisheries nationally. Little is known about the mechanism of these entanglements. To protect these animals as required by the Marine Mammal Protection Act (MMPA), fishery managers use time area closures and/or permit the use of acoustic deterrent devices (pingers) to reduce bycatch. To understand how an entanglement occurs and how the pingers function, one must understand the underwater behavior of the animals. Studies of captive porpoises, and using mounted instrumentation on released animals, have shown some aspects of porpoise life history but may or may not reveal natural behavior. A research program was initiated to determine the feasibility of passively tracking wild harbor porpoise using their echolocation signals (clicks) and observing their underwater movements in three dimensions. Additionally, details of these movements during feeding, traveling, and social events may provide an insight into their behavior that makes them prone to entanglement. Preliminary results demonstrate that the underwater tracking of harbor porpoises can be accomplished using this technique and several aspects of the animal's natural behavior and communication are presented. A next generation autonomous system, to be deployed near actively fishing sink gillnets being developed, is described. [Work supported by NMFS and the URI/NOAA CMER Program.]

1:30

5pAB3. Passive acoustic tracking of echolocating harbor porpoises (*Phocoena phocoena*). II. Methodology and results. James H. Miller, Robert Gampert (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882), David C. Potter (National Marine Fisheries Service, Woods Hole, MA 02543), and Khosrow Lashkari (MBARI, Moss Landing, CA 95039)

Harbor porpoises (*Phocoena phocoena*) generate ultrasonic echolocation signals in the 110–150 kHz frequency range and social calls below 20 kHz. The high-frequency ''clicks'' can be used to acoustically localize and track these animals in three dimensions at ranges within 500 m. In addition, the lower-frequency social calls may be used for longer-range localization. This presentation describes the techniques and methods for passively tracking harbor porpoises using these echolocation clicks and social calls. The approach is based on a time-difference-of-arrival (TDOA) direction finding and line-of-position (LOP) fixing from two arrays spaced 60-m apart. Each bottom-mounted array consists of four hydrophones in a 1-m tetrahedral configuration cabled 300 m to shore. Porpoise signal characteristics such as frequency, bandwidth, pulse rate, source level, and beamwidth and their effects on tracking are reviewed. Various components of the tracking system such as the hydrophone array, signal-conditioning electronics, underwater cable, recording, and postprocessing systems, and the tracking algorithms are described. Explicit localization solutions are presented for this array geometry and performance bounds are given. Tracking results are presented using both the low- and high-frequency signal types received from various ranges and bearings during a one-week experiment in the Gulf of Maine. [Work supported by NMFS and the URI/NOAA CMER Program.]

1:45

5pAB4. Inexpensive passive acoustic animal tracking system. Sean A. Hayes, Daniel P. Costa, Donald Croll (Bio. Dept. and IMS, Univ. of California, Santa Cruz, CA 95064, hayes@biology.ucsc.edu), David K. Mellinger (Monterey Bay Aquarium Res. Inst., Moss Landing, CA 95039), and J. Fabrizio Borsani (Tethys Res. Inst., Milan, Italy)

Marine bioacoustic research is challenged by economic, logistical, and technical barriers. A multibuoy recording system has been designed. Each buoy contains a DAT recorder (SONY D-8; 9 Hz–22 kHz) with a hydrophone (Hi-Tech HTI-SSQ-41B; 10 Hz–30 kHz) connected to one channel for collecting passive acoustic data. The second channel records time-synchronizing VHF radio signals, for computing time-of-arrival differences between buoys. A GPS data logger documents buoy position. This package is encased in a PVC spar-buoy housing. During analysis, cross correlation and peak picking time synchronize the separate recordings. Cross-correlation or waveform measurement is used to determine time-of-arrival differences of vocalizations between buoys. A hyperbolic least-squares fixing algorithm is used to calculate locations of vocalizing animals. In a calibration test, three buoys were deployed in triangle formation (1.8-km spacing). Light bulb implosions were localized with an accuracy of 0.20 km. The buoys are far less expensive (approximately $1900 each) than systems requiring towed arrays, multichannel signal conditioning/

recording, and large tow ships, or fixed bottom-mounted arrays. Buoys are deployable wherever animals are located, in sea state conditions through Beaufort 5. These buoys have recorded eight marine mammal species. [Work supported by ONR, ATOC, AMNH, Friends of LML, and Myers trust fund.]

2:00

5pAB5. A feasibility field study of monitoring blue whales using the Pt. Sur Ocean Acoustic Observatory. Ching-Sang Chiu, Curtis A. Collins, Carl A. Hager, Christopher W. Miller, Therese C. Moore, Michael R. Rocheleau (Dept. of Oceanogr., Naval Postgrad. School, Monterey, CA 93943), Khosrow Lashkari (Monterey Bay Aquarium Res. Inst., Monterey, CA), and Sean Hayes (Univ. of California, Santa Cruz, CA)

In the summer of 1997, two three-day experiments were conducted to test the feasibility of acoustically detecting, classifying, localizing, and tracking blue whales at long ranges using a former SOSUS listening array located at the Naval Postgraduate School Ocean Acoustic Observatory (OAO) at Pt. Sur, California. During each experiment, full-array data were archived continuously at the OAO. In concert with the shore-based acoustic monitoring, an aircraft was assigned to locate blue whales in the Monterey Bay National Marine Sanctuary and to direct a research vessel to a whale site. The research vessel was manned with observers and instrumented with a towed hydrophone array to ground-truth the locations of the blue whales and classify the vocalized near-field signals. These shipboard measurements were required to provide a means to separate the source signal characteristics from the multipath signatures for the calibration and validation of broadband, model-based localization methods. In this presentation, an overview of this feasibility field study is provided. Initial experimental as well as modeling results are presented, which include assessments of the predictability (i.e., variability) of the vocalized sound and the uniqueness of the location-dependent multipath structure, both are fundamental to the applicability of model-based algorithms. [Research supported by ONR.]

2:15

5pAB6. Acoustical localization of dolphins in shallow water. Roland Aubauer, Marc O. Lammers, and Whitlow W. L. Au (Hawaii Inst. of Marine Biol., Univ. of Hawaii, P.O. Box 1106, Kailua, HI 96734, raubauer@soest.hawaii.edu)

Previous attempts at localization of cetaceans have generally used multiple hydrophone arrays and multichannel recording systems. In this paper a low-budget localization technique using only one hydrophone will be described. The time delays of the signals traveling on the surface and bottom reflection paths to the hydrophone, relative to the direct signal, are used to calculate the distance and the depth of a phonating animal. Only two additional measures, the depth of the bottom and hydrophone, have to be taken. The method requires relatively shallow waters and a flat ground surface. The received signals are digitized directly into a laptop computer via a PCMCIA A/D converter card with a sampling rate of 260 ksamples/s. Hawaiian spinner dolphins (*Stenella langirostris*) were localized by echolocation and burst pulse click trains over different bottom substrates. The method is most valuable for examining source levels of dolphin clicks in the wild and could lead to a better understanding of the nature and use of click trains by dolphins. [Work supported by Deutsche Forschungsgemeinschaft and Office of Naval Research.]

2:30

5pAB7. Acoustic census of sperm whales in the eastern temperate North Pacific. Jay Barlow and Barbara Taylor (Southwest Fisheries Sci. Ctr., P.O. Box 271, La Jolla, CA 92038, jayb@pooh.ucsd.edu)

An NOAA/ONR acoustic line-transect survey was conducted in spring 1997 to estimate the abundance of sperm whales in the eastern temperate North Pacific using a 52-m research vessel. An ITI hydrophone array was towed at 100-m depth as the vessel surveyed approximately 25 000 km of predetermined transect lines at 8 kn. Acoustic signals received by a pair of hydrophones were digitized and sent up to the ship via a coaxial tow cable. Bearing angles to the locations of sperm whales were calculated in real time based on differences in signal arrival times between the two hydrophones, and whale locations were determined by the convergence of bearing lines as the ship continued along its course. The ship was maneuvered to resolve the left/right ambiguity in whale locations and, during daylight hours, was directed to this estimated location to obtain visual estimates of group size when the whales surfaced. A total of 102 distinct groups of sperm whales were acoustically detected with sufficient amplitude to estimate bearing lines. ''Slow clicks'' (with a period of 4–8 s) were detected at ranges up to 20 nmi, whereas the more typical sperm whale clicks (with a period of 0.5–1.0 s) were typically heard at less than 5 nmi.

2:45

5pAB8. Acoustical models and measurements of individual Antarctic krill. Duncan E. McGehee (Tracor Appl. Sci., 4669 Murphy Canyon Rd., Ste. 102, San Diego, CA 92123), Richard L. O'Driscoll (Univ. of Otago, Dunedin, New Zealand), and Linda V. Martin Traykovski (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543-1049)

Acoustic surveys are commonly used to estimate the biomass of Antarctic krill (*Euphausia superba*). Accurate biomass estimates depend upon knowing the acoustic backscattering properties of the animals. A generic scattering model for individual *E. superba* was developed and tested against laboratory measurements. The model was based on morphometric measurements of live krill and employed a simplified form of the distorted wave Born approximation (DWBA) method. The model included the effects of animal length and orientation. Backscattering measurements of live individual Antarctic krill were made at 120 kHz as part of the August, 1995 Bioacoustical Oceanography Workshop in Santa Cruz, CA, jointly sponsored by the National Science Foundation and the Office of Naval Research. Animals were suspended one by one in an acoustic beam, with substantial freedom to move and thus with more or less random orientation. One-thousand echoes were collected per animal, and target strength was calculated for each echo. Behavior was recorded on video. The video data were used to estimate three-dimensional orientation for five of the animals, thus giving their target strengths as functions of orientation. The measurements supported the major features of the scattering model.

3:00

5pAB9. *In-situ* acousto-optical imaging of zooplankton with the OASIS system. Jules S. Jaffe, Mark D. Ohman, and Alex De Robertis (Scripps Inst. of Oceanogr., La Jolla, CA 92093)

A new instrument has recently been developed: OASIS (optical-acoustical submersible imaging system), for recording *in-situ* acoustic target strengths (TS) of crustacean zooplankton with concurrent optical imaging of the animals. The system consists of a three-dimensional acoustic imaging system (Fish TV) and a sensitive optical CCD camera with a red-filtered strobe. The sonar triggers the acquisition of an optical image when it detects the presence of a significant target in the precise location where the camera and strobe are co-registered. Acoustic TS can then be related to the optical image which permits the identification of the animal and its 3-D orientation. The system was deployed in Saanich Inlet, BC, Canada in the summers of '86 and '87. Target strengths of many acoustic reflectors were recorded in parallel with the optical images triggered by the presence of an animal in the correct location of the sonar system. Inspection of the optical images revealed that the primary zooplankton taxa at the site were euphausiids [esp. *Euphausia pacifica*] and [gammarid] amphipods [esp. *Cyphocaris challengeri*] and permitted an exact correlation of TS and taxa. Extrapolation of this information to many other acoustic targets, imaged by the Fish TV system will allow more precise interpretation of both abundance and behavior.

3:15

5pAB10. Incoherent localization of targets for a three-dimensional multibeam sonar imaging system. Jules S. Jaffe (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238)

One advantage that three-dimensional multibeam imaging systems have over more conventional systems is that they have the potential to track many targets that are concurrently in their field of view. As such, the accurate estimation of both position and target strength are important data which can be derived from the echo responses of an active multibeam imaging system. An algorithm has recently been developed which can be used to obtain higher positional estimates for targets than available from the preformed beams of a given multibeam system (Fish TV). The method refines a least-squares residual of the amplitudes of a set of predicted data versus a set of observed data by adjusting the parameters of a model. The parameters are the position and target strength of the target. The least-squares norm results in a maximum likelihood estimator for position and target strength in high signal-to-noise. Two fortunate aspects of the algorithm are that it takes into account the geometry of the sonar system and also uses only the amplitude data. The special collection geometry and acquisition method make these features important.

FRIDAY AFTERNOON, 5 DECEMBER 1997 CALIFORNIA ROOM, 1:00 TO 4:45 P.M.

Session 5pUW

Underwater Acoustics: Reverberation and Scattering

Xavier Zabalgogeazcoa, Chair
Science Applications International Corporation, Mail Code C4A, 10260 Campus Point Drive, San Diego, California 92121

Contributed Papers

1:00

5pUW1. High-frequency surface forward scattering and its relation to the directional wave spectrum: Model and data comparisons employing the Kirchhoff approximation. Peter H. Dahl (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98195)

Measurements of the spatial coherence of high-frequency (30-kHz) sound forward scattered from the sea surface are discussed, along with a simple interpretive model based on the Kirchhoff approximation. The data are from an experiment conducted off south Florida in waters 25-m deep. For the model, the bistatic cross section is first computed via the Kirchhoff integral, and a marginal density function for intensity versus horizontal arrival angle generated; this function is then Fourier transformed to give a model for the horizontal spatial coherence. The Kirchhoff integral requires a representation of the two-dimensional surface wave correlation function, which is itself derived from estimates of the directional wave spectrum. The latter were measured with a directional wave buoy that operated in the immediate vicinity of the acoustic measurements. A model [Donelan *et al.*, Philos. Trans. R. Soc. London, Ser. A **315**, 509–562 (1985)] is used to fill in for the necessary directional *spreading* information. Sensitivities of horizontal spatial coherence to array acquisition geometry, wind speed, surface wave directional properties and mean-square slope, are discussed. [Work supported by ONR Code 321OA.]

1:15

5pUW2. A simple approach for modeling acoustic scattering from rocks and sea shells lying on the ocean floor. C. Feuillade and R. W. Meredith (Naval Res. Lab., Stennis Space Center, MS 39529-5004)

Acoustic reverberation in the ocean is due to collective scattering from a large number and diverse range of distributed objects. In this work, a model for predicting the sonar target strengths of an important class of clutter objects, i.e., rocks and sea shells lying on the ocean floor, is presented. The model incorporates a simplified approach, in which rocks are represented by either rigid movable spheres or elastic spheres, a technique which has been previously used to model aqueous suspensions of sand over the frequency range of interest. Sea shells are represented by water-filled spherical shells. The model predicts scattering strength values for areas of the ocean floor covered by randomized arrangements of rocks and shells, using sizes, numbers, geologic type, and material compositions typically found on the bottom. The scattering as a function of azimuth is determined by coherent addition, and allows both monostatic and bistatic geometries to be investigated. [Work supported by ONR/NRL]

1:30

5pUW3. Multiple scattering in waveguides with rough interfaces. David H. Berman (Dept. of Phys. and Astron., Univ. of Iowa, Iowa City, IA 52245, dhb@astro.physics.uiowa.edu)

Schmidt and Kuperman [J. Acoust. Soc. Am. **97**, 2199–2209 (1995)] show that it is possible to treat multiple scattering in media with elastic and fluid layers separated by rough interfaces. Their work, based on a ''smoothing'' approximation, shows that the coherent (mean) field obeys certain effective boundary conditions. These boundary conditions account for some of the multiple interactions between interfaces. Kuperman and Schmidt also compute the second moment of the scattered field. However, their computations of the second moment include multiple scattering only insofar as the fields which are scattered are the mean fields which depend on the effective boundary conditions. Recent work [Sanchez-Gil *et al.*, Phys. Rev. B **50**, 15353–15368] shows that to account for enhanced back-scattering from roughness overlying a waveguide structure, multiple scattering theory for the incoherent field is required, beyond the smoothing approximation. This work will examine consequences of their multiple scattering formalism for fields within a waveguide and identify the problems that make the interior problem distinct from that of determining the differential cross section, i.e., the exterior problem.

1:45

5pUW4. Small-scale roughness and the parabolic equation. Daniel Wurmser and Robert F. Gragg (Code 7144, Naval Res. Lab., Washington, DC 20375-5350, wurmser@abyss.nrl.navy.mil)

When the Foldy–Wouthuysen transformation is used to derive the parabolic equation (PE) for a range-dependent environment, a new physical effect associated with small-scale fluctuations of the field is predicted [Wurmser *et al.*, J. Acoust. Soc. Am. **101**, 1309–1327 (1997)]. In this presentation, the formalism is applied to the problem of scattering from a rough surface, which is treated as a type of range dependence. The theory predicts the effective impedance boundary conditions proportional to the curvature of the surface, essentially unifying the Biot–Tolstoy theory of scattering from a bossed surface with other modern scattering and propagation theories. The result has been obtained for the two-fluid, Neumann

and Dirichlet boundary conditions. The related physical effect is associated with the small-scale cutoff of the field, implying that the wave cannot resolve a surface structure smaller than about 1/12 of a wavelength. The theory also predicts related effects; for example, roughness-induced interface waves for the Neumann and two-fluid boundary conditions. Since the effect on the scattered field is cumulative, its incorporation should improve PE modeling of propagation in waveguides with rough boundaries. [Work supported by ONR.]

2:00

5pUW5. Beyond direct path seabottom scatter. Andrew Rogers and Peter Neumann (Planning Systems, Inc., 7923 Jones Branch Dr., McLean, VA 22102, arogers@plansys.com)

A great deal of reverberation data collected and processed for scattering strength is truncated such that angles containing energy from fathometer and hybrid eigenray paths are disregarded. This practice has been a result of the common use of plane-wave direct path scattering models, which inherently cannot compute these events. The recent development of the acoustic scattering model SCARAB (scattering reverberation and backscatter) has provided a tool for investigating the information contained in this low grazing angle scatter. Data taken during the Critical Sea Test (CST) initiative has been processed using traditional direct path methods retaining data at scattering angles well below that where fathometer and hybrid path contamination occur. Model to data comparisons have shown excellent agreement, providing new insight into the dominant mechanisms for scattering from the seabottom. Utilization of the low grazing angle fathometer and hybrid paths is found to be critical in defining the physical control mechanisms of seabottom scatter at all angles. [Work supported by ONR.]

2:15

5pUW6. Physics-based reverberation and scatter modeling. Andrew Rogers and Peter Neumann (Planning Systems, Inc., 7923 Jones Branch Dr., McLean, VA 22102, arogers@plansys.com)

The SCARAB (scattering reverberation and backscatter) model takes a non-plane-wave, physics-based approach to the three-dimensional problem of ocean reverberation and seabottom scatter. Using a ray propagation model, coupled with a geoacoustic description of the sediment and physics-based interface and volume sediment scattering kernels, an ocean environment's reverberative, and scattering capability is found for given source and receiver operating characteristics. Advantages of this approach over conventional direct path methods are that received levels are modeled as if they were experimentally measured functions of time. Therefore, subsequent processing for scattering level can follow the same procedures as are used for experimental data. This has proven to be of great importance when modeling low grazing angle bottom scatter contaminated with fathometer and hybrid path contributions. A description of the model is given and model-to-data comparisons are shown for data taken during the Critical Sea Test (CST) experiments. [Work supported by ONR.]

2:30

5pUW7. Multiple scattering owing to sediment volume inhomogeneities. Dajun Tang (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

Multiple scattering owing to strong sediment volume inhomogeneities is discussed using the formally averaged approach, which was originated from the quantum field theory and has been extensively applied to wave propagation in random media. Random spatial variations of sediment density as well as sound speed are considered. While it is well known that the sound-speed variations result in a renormalized complex wave number, when the coherent field is concerned it is found that the density variations can also be renormalized, but they manifest themselves as a formal density gradient in the wave equation. This can be identified as a new loss mechanism of the coherent field. When both sound speed and density variations are included, the cross correlation between sound speed and density is an important quantity to measure since it will influence the scattering intensity to a non-negligible extent. Numerical examples will be presented and discussed. [Work supported by ONR Code 321OA.]

2:45

5pUW8. Modeling high-frequency reverberation data from coastal waters. Stanley A. Chin-Bing, Steve Stanic (Naval Res. Lab., Stennis Space Center, MS 39529-5004), and Joseph E. Murphy (Univ. of New Orleans, New Orleans, LA 70129)

High-frequency (40–180 kHz) acoustic reverberation data have been taken from a region 26 miles south of Panama City, Florida. The data sets were obtained from a series of bottom scattering measurements taken in an area where the bottom was covered with coarse sand and shell fragments. Two different high-frequency scattering models designed for high-frequency reverberation predictions had previously failed to match the data. These data have been reexamined using highly accurate ocean acoustic computer models based on the finite-element solution to the Helmholtz wave equation. While these models were designed for lower frequency predictions, the availability of supercomputers allowed us to apply these models at these high frequencies. In this study, new comparisons are given to the data and to the previous modeling predictions. [Work supported by ONR/NRL and a High Performance Computing DoD grant.]

3:00–3:15 Break

3:15

5pUW9. Broadband bottom scattering, loss, and reverberation measurements near Elba Island. Charles W. Holland (SACLANT Undersea Res. Ctr., 19138 La Spezia, Italy)

The prediction of low- to midfrequency shallow-water reverberation is still in its infancy. This is due, in large part, to the unpredictability of the bottom-scattering kernel, and particularly, its dependence upon frequency. The bottom-scattering kernel is difficult to measure directly and even more problematic to extract from the reverberation. An experiment was conducted in a complex shallow-water area to test novel coherent and incoherent direct-path bottom-scatter measurement techniques (400–4000 Hz). Advantages and disadvantages of each technique are discussed. Analyses of the bottom-scatter data in conjunction with collocated bottom loss and high-resolution seismic data shed light on the underlying mechanisms (e.g., interface roughness, sub-bottom volume inhomogeneities, or basement roughness) that give rise to the scattering. It is hypothesized that the frequency dependence of the scattering kernel can be linked to the dominant scattering mechanism.

3:30

5pUW10. Coherent reflection loss from a Pierson–Moskowitz sea surface using the NLSSA. Shira L. Broschat (School of Elec. Eng. and Comput. Sci., Washington State Univ., P.O. Box 642752, Pullman, WA 99164-2752)

Numerical results for the incoherent bistatic scattering strength for the nonlocal small-slope approximation (NLSSA) for scattering from rough surfaces [A. G. Voronovich, Waves Random Media **6**, 151–167 (1996)], using a Gaussian spectrum, were recently reported [S. L. Broschat and E. I. Thorsos, J. Acoust. Soc. Am. **100**, 2702 (1996)]. The cases considered agreed well with exact integral equation results and demonstrated the potential of the NLSSA for accurate modeling of rough surface scattering. In this paper, results for the coherent reflection loss for one-dimensional surfaces are presented. Coherent loss is of particular importance in long-range acoustic propagation when the cumulative effect of multiple interactions with the surface can be significant. The lowest-order NLSSA reflection coefficient is derived, and numerical results are shown for the reflection loss at low grazing angles, using a Pierson–Moskowitz sea surface spectrum. The numerical results are compared with those of other approximate methods and agree well with exact integral equation results. [Work supported by ONR.]

3:45

5pUW11. Acoustic propagation across a roughened fluid–fluid interface. Joseph L. Lopes, Iris C. Paustian, and Raymond Lim (Coastal Systems Station, Dalghren Div., Code R21, 6703 West Hwy. 98, Panama City, FL 32407-7001)

A laboratory experiment was conducted to investigate the effects of small-scale random roughness on acoustic propagation across a fluid–fluid interface. This experiment utilized an acoustic source mounted onto a movable platform, several hydrophones, and two immiscible fluids (vegetable oil and glycerin). The effects of roughness on propagation across the interface were determined by floating small polystyrene beads at the interface between these fluids. Data were collected as functions of frequency, grazing angle, and varying degrees of roughness (bead size and with/without beads). The goal of this experiment was twofold: (1) to test roughness diffraction as a mechanism for enhanced acoustic transmission at grazing angles shallower than critical, and (2) to assess the temporal and spatial coherence of the transmitted signal. Analyzed data of this work will be presented and compared to results of propagation into underwater sediments reported in the literature. [Work supported by ONR.]

4:00

5pUW12. Accuracy of the narrow-band approximation for the synthesis of low-frequency time series in ocean waveguides. Kevin D. LePage (BBN Systems and Technol., 70 Fawcett St., Cambridge, MA 02138)

The expansion of modal wave numbers about a center frequency is considered for the efficient synthesis of low-frequency time series of acoustic pulses propagated at long ranges in the ocean. For absolutely bandlimited signals, the complex envelope is obtained directly in terms of the Fresnel integral, allowing very rapid computation of time series at long range. The quality of the approximation is evaluated for both shallow and deep water scenarios as a function of center frequency and bandwidth. Results show that time series with bandwidths of up to approximately half the center frequency may be reliably synthesized using these types of techniques for center frequencies above 20 Hz.

4:15

5pUW13. Wall pressure and acoustic modeling from boundary-layer bursts and sweeps. James M. Witting (Appl. Res. Lab., Penn State Univ., State College, PA 16804)

Explicit treatment of discrete, transitory, momentum-transporting features in boundary layers (bursts and sweeps) can generate acoustic models [e.g., J. M. Witting, ''A spectral model of pressure fluctuations at a rigid wall bounding an incompressible fluid, based on turbulent structures in the boundary layer,'' Noise Control Eng. J. **26**, 28–43 (1986)]. This presentation addresses the following topics: (1) the relationships between the properties of the microscopic structures (such as their sizes, strengths, and durations) and the details of the wall-pressure spectrum; (2) estimating actual microscopic properties from measured data, either directly or from empirical models that have been compared to measured data; (3) extensions to more complex geometries; and (4) the role of the boundary-layer structures as direct sources of sound in the fluid.

4:30

5pUW14. Effects of cross correlation between different types of medium perturbations on seabed scattering. Anatoliy N. Ivakin (Andreev Acoust. Inst., Shvernika 4, Moscow 117036, Russia, aniva@glasnet.ru)

Sound scattering from the seabed can be generally attributed to two major mechanisms connected with seabed roughness and volume inhomogeneities. Each of these two mechanisms in turn can contain different components. For unconsolidated (fluid) sediments, volume inhomogeneities are due to spatial fluctuations of two different parameters, the density and compressibility, or the density and compressional wave speed, with respect to their mean values. For elastic bottoms, one more fluctuating parameter, shear wave speed, is to be considered. The roughness component can be attributed to the water–sediment interface and to various internal interfaces in the seabed medium as well. In this paper, the effects of cross correlations between the roughness of different interfaces and between the volume fluctuations of different parameters, as well as possible volume–roughness correlations, are examined using a first-order perturbation model. Expressions for the bottom scattering cross section involve corresponding cross-correlation and cross-spectral matrices of seabed medium perturbations. Frequency-angular dependencies of the scattering strength in monostatic and bistatic cases are analyzed for various types of seabed stratification and correlated irregularities. Comparisons are presented for the cases of strongly correlated, anticorrelated, and partially correlated perturbations. Examples are given where the cross-correlation effects are important and should be taken into account in development of methods for inversions of seabed properties. [Work supported by ONR.]